CENTER FOR STATISTICAL RESEARCH AND METHODOLOGY
FY 2016 FIRST & SECOND QUARTERS REPORT

–October 2015 through March 2016 –

COLLABORATION

DECENNIAL DIRECTORATE

Decennial Management Division/Decennial Statistical Studies Division/American Community Survey Office (Sponsors)

Project Number	Project Title	FTEs
6650B23	Redesigning Field Operations	1.32
6750B01	Administrative Records Data	3.87
	A. Decennial Record Linkage	
	B. Coverage Measurement Research	
	C. Using 2010 Census Coverage Measurement Data to Compare Nonresponse Follow-up Proxy Responses with Administrative Records	
	D. Record Linkage Error-Rate Estimation Methods	
	E. Supplementing and Supporting Non-Response with Administrative Records	
	F. Identifying "Good" Administrative Records for 2020 Census NRFU Curtailment Targeting	
	G. Evaluation of Response Error Using Administrative Records	
	H. Special Census: Disclosure Avoidance in Group Quarters	
	I. 2020 Unduplication Research	
6350B02	Address Canvassing in Field	1.59
	A. Master Address File (MAF) Error Model and Quality Assessment	
	B. Development of Block Tracking Database	
	C. Detection of Map Changes	
6550B01	Data Coding, Editing, and Imputation	0.15
6250B07	Policy	0.25
	A. Privacy and Confidentiality for the 2020 Census	
6385B70	American Community Survey (ACS)	2.72
	A. ACS Applications for Time Series Methods	
	B. ACS Imputation Research and Development	
	C. Data Analysis of ACS CATI-CAPI Contact History	
	D. Assessing Uncertainty in ACS Ranking Tables	
	E. Confidence Intervals for Proportions in ACS Data	
	F. Mode-Based Imputation in ACS	
	G. Voting Rights Section 203 Model Evaluation and Enhancements Towards Mid-Decadal Determinations	

DEMOGRAPHIC DIRECTORATE

Demographic Statistical Methods Division (Sponsor)

Project Number	Project Title	FTEs
TBA	Demographic Statistical Division Special Projects	TBA
	A. Special Project on Weighting and Estimation	
	B. Weighted Estimating Equations with Response Propensities in Terms of Covariates Observed Only for Responders	

Demographic Surveys Division (Sponsor)

Project	Project	
Number	Title	FTEs
0906/1444X00	Demographic Surveys Division Special Projects	0.70
	A. Data Integration	

Population Division (Sponsor)

Project	Project	
Number	Title	FTEs
TBA	Population Division Projects	TBA
	A. Introductory Sampling Workshop	

Social, Economic, and Housing Statistics Division (Sponsor)

Project	Project	
Number	Title	FTEs
7165016	Social, Economic, and Housing Statistics Division Small Area Estimation Projects	2.38
	A. Research for Small Area Income and Poverty Estimates (SAIPE)	
	B. Small Area Health Insurance Estimates (SAHIE)	
	C. Sub County Estimates of Poverty from Multi-year ACS Data	

ECONOMIC DIRECTORATE

Project	Project	
Number	Title	FTEs
1183X01	Economic Statistical Collection	0.25
1001X00	Economic Monthly/Retail	0.06
	A. Research on Imputation Methodology for the Monthly Wholesale Trade Survey	
	B. Use of Big Data for Retail Sales	
2220B10	Economic Census/Survey Engineering: Time Series Research; Economic Missing Data/Product Line Data; Development/SAS	3.25
	A. Seasonal Adjustment Support	
	B. Seasonal Adjustment Software Development and Evaluation	
	C. Research on Seasonal Time Series - Modeling and Adjustment Issues	
	D. Supporting Documentation and Software for X-13ARIMA-SEATS	
	E. Missing Data Adjustment Methods for Product Data in the Economic Census	
7103012	2012 Commodity Flow Survey	0.03
	A. Commodity Flow Survey	
TBA	Investigation of Alternative Methods for Resolving Balance Complex Failures in StEPS	TBA
	A. Investigation of Alternative Methods for Resolving Balance Complex Failures in StEPS	

RESEARCH AND METHODOLOGY DIRECTORATE

Center for Economic Studies (Sponsor)

Project	Project	
Number	Title	FTEs
TBA	Business Dynamics Statistics—Export Weighting Issue	TBA
	A. Business Dynamics Statistics—Export Weighting Issue	

CENSUS BUREAU

Project Number	Project Title	FTEs
7236045	National Survey of Drug Use & Health	0.06
	A. National Survey of Drug Use & Health	
0331000	Program Division Overhead	9.00
	A. Center Leadership and Support	
	B. Research Computing	

GENERAL RESEARCH AND SUPPORT

Project Number	Project Title	FTEs
0331000	General Research and Support	26.05
0925000	General Research	8.28
MISSING DATA, EDIT, AND IMPUTATION		
	A. Editing	
	B. Editing and Imputation	
RECORD LINKAGE		
	A. Disclosure Avoidance for Microdata	
	B. Record Linkage and Analytic Uses of Administrative Lists	
	C. Modeling, Analysis and Quality of Data	
	D. R Users Group	
SMALL AREA ESTIMATION		
	A. Small Area Methods with Misspecification	
	B. Coverage Properties of Confidence Intervals for Proportions in Complex Surveys	
	C. Small Area Estimates of Disability	
	D. Using ACS Estimates to Improve Estimates from Smaller Surveys via Bivariate Small Estimation Models	
	E. Multivariate Fay-Herriot Hierarchical Bayesian Estimation of Small Area Means under Functional Measurement Error	
	F. Smoothing Design Effects for Small Sample Areas	
SURVEY SAMPLING-ESTIMATION AND MODELING		
	A. Household Survey Design and Estimation	
	B. Sampling and Estimation Methodology: Economic Surveys	
	C. The Ranking Project: Methodology Development and Evaluation	
	D. Sampling and Apportionment	
	E. Interviewer-Respondent Interactions: Gaining Cooperation	
TIME SERIES AND SEASONAL ADJUSTMENT		
	A. Seasonal Adjustment	
	B. Time Series Analysis	
	C. Time Series Model Development	
EXPERIMENTATION AND STATISTICAL MODELING		
	A. Design and Analysis of Embedded Experiments	
	B. Synthetic Survey and Processing Experiments	
	C. Multivariate Nonparametric Tolerance Regions	
	D. Master Address File (MAF) Research—Developing a Generalized Regression Model for Count Data	
	E. Modeling the Causal Effects of Field Representative Actions and Strategies	
	F. Development of a Bivariate Distribution for Count Data where Data Dispersion is Present	
	G. Developing a Flexible Stochastic Process for Significantly Dispersed Count Data	
SIMULATION AND STATISTICAL MODELING		
	A. Development and Evaluation of Methodology for Statistical Disclosure Control	
SUMMER AT CENSUS		
RESEARCH SUPPORT AND ASSISTANCE		

PUBLICATIONS

- Journal Articles, Publications
- Books/Book Chapters
- Proceedings Papers
- Center for Statistical Research & Methodology Research Reports
- Center for Statistical Research & Methodology Studies

TALKS AND PRESENTATIONS

CENTER FOR STATISTICAL RESEARCH & METHODOLOGY SEMINAR SERIES

PERSONNEL ITEMS

- Honors/Awards/Special Recognition
- Significant Service to Profession
- Personnel Notes

1. COLLABORATION

1.1 REDESIGNING FIELD OPERATIONS (Decennial Project 6650B23)

1.2 ADMINISTRATIVE RECORDS DATA (Decennial Project 6750B01)

A. Decennial Record Linkage

Description: Under this project, staff will provide advice, develop computer matching systems, and develop and perform analytic methods for adjusting statistical analyses for computer matching error with a decennial focus.

Highlights: Staff provided extensive comments to the Decennial Statistical Studies Division (DSSD) related to background on Decennial record linkage methods and production software. The background covered what name and address standardization software were available and why they were crucial to Decennial processing and research. The background also covered what software (most funded by DSSD) had been written for Decennial processing (the SRD 1-1 matcher used in 1990, 2000, and 2010 and *BigMatch* used in 2010). There are at least seven variants of the software that have been written for DSSD but not used for production and other variants for projects in the Economic Directorate.

Staff: William Winkler (x34729), Emanuel Ben-David

B. Coverage Measurement Research

Description: Staff members conduct research on model-based small area estimation of census coverage, and they consult and collaborate on modeling census coverage measurement (CCM).

Highlights: No significant updates these quarters.

Staff: Jerry Maples (x32873), Ryan Janicki, Eric Slud

C. Using 2010 Census Coverage Measurement Data to Compare Nonresponse Follow-up Proxy Responses with Administrative Records

Description: Research in preparation for the 2020 Census Nonresponse Follow-up (NRFU) investigates employing different contact strategies combined with the use of administrative records (AR) files in different ways in order to reduce the cost of the operation while maintaining data quality. Regardless of the contact strategy, the question arises as to whether the proxy responses are more accurate than ARs available for the NRFU housing units (HUs). The goal of this study is to use the results of the 2010 Census Coverage Measurement Program (CCM) to compare the accuracy of proxy responses for 2010 Census NRFU housing units in the CCM sample with the accuracy of the ARs available for the housing units.

Highlights: A draft research report is undergoing internal Census Bureau review.

Staff: Mary Mulry (x31759)

D. Record Linkage Error-Rate Estimation Methods

Description: This project develops methods for estimating false-match and false-nonmatch rates without training data and with exceptionally small amounts of judiciously chosen training data. It also develops methods/software for adjusting statistical analyses of merged files when there is linkage error.

Highlights: Previously six staff members worked on automatic error-rate estimation for record linkage for more than fifteen months. The staff included individuals from CSRM, the Decennial Statistical Studies Division (DSSD), and the Center for Administrative Records Research and Applications (CARRA). More recently, this project has been on hold.

Staff: William E. Winkler (x34729), Emanuel Ben-David, Tom Mule (DSSD), Lynn Imel (DSSD), Mary Layne (CARRA)

E. Supplementing and Supporting Non-Response with Administrative Records

Description: This project researches how to use administrative records in the planning, preparation, and implementation of nonresponse follow-up to significantly reduce decennial census cost while maintaining quality. The project is coordinated by one of the 2020 Census Integrated Project Teams.

Highlights: Staff continued to analyze the results of models (primarily stepwise logistic regression) with Census Unedited File (CUF) household size as the dependent variable. The models used two main data files: a file of Nonresponse Follow-up (NRFU) IDs for Maricopa County and a national file of NRFU IDs. The models mostly focused on the subset of NRFU IDs where the 2009 IRS 1040 return was present and the undeliverable as addressed (UAA) flag was blank. Fitting a separate logistic regression for each value of the top-coded IRS 1040 household count was needed for the predicted household size (household size with the highest predicted probability) distribution to be reasonably similar to the actual CUF distribution. Even the model with separate logistic regressions did not seem to perform much better than simply using the IRS 1040 household count. Staff also fit models (on both the Maricopa data file and the national data file) on a random 5% subsample and scored the models on the entire file. The fitting and scoring were both done in four separate pieces (UAA flag blank and 2009 IRS

1040 present, UAA flag blank and 2009 IRS 1040 absent but 2008 IRS 1040 present, UAA flag blank and both 2008 and 2009 IRS 1040 absent, UAA flag nonblank) and the results from the four pieces were combined into a single output file. The scored output (including predicted probabilities, predicted household size, and input model variables) was provided to staff in the Decennial Statistical Studies Division (DSSD). DSSD used the predicted probabilities to incorporate constraints on the expected value of household size (both housing unit and aggregate level constraints were considered) into models for using administrative records for occupied housing units. Adding constraints on expected household size does show some promise for helping to maintain the overall population count.

In addition, staff analyzed data collected in the 2015 Evaluation Follow-up (EFU) conducted after the Nonresponse Follow-up (NRFU) operations in the 2015 Census Test in Maricopa County, AZ. The 2015 EFU collected data for a comparison of the accuracy of occupancy status and responses for 2015 NRFU housing units (HUs) with the accuracy of the administrative records (AR) available for the housing unit using data from the 2015 EFU. The EFU selected a sample of 4,098 HUs in NRFU where there was at least one of nine types of discrepancies between the NRFU and AR information for the HU and conducted interviews July 20 to August 14, 2015. The result that EFU had higher agreement with NRFU than AR for 43 percent of the cases where the count was differing led to the recommendation to include an additional mailing for the AR units in the future. Ambiguities in the Undeliverable As Addressed (UAA) forms received from the U.S. Postal Service (USPS) regarding HUs receiving a status of vacant by ARs but a status of occupied by NRFU led to the recommendation to do qualitative research with USPS during the 2016 Census Test. A draft report is undergoing internal Census Bureau review.

Staff: Michael Ikeda (x31756), Mary Mulry

F. Identifying "Good" Administrative Records for 2020 Census NRFU Curtailment Targeting

Description: As part of the Census 2020 Administrative Records Modeling Team, staff are researching scenarios of nonresponse follow-up (NRFU) contact strategies and utilization of administrative records data. Staff want to identify scenarios that have reduction in NRFU workloads while still maintaining good census coverage. Staff are researching identification of "good" administrative records via models of the match between Census and administrative records person/address assignments for use in deciding which NRFU households to continue to contact and which to primary allocate. Staff are exploring various models, methods, and classification rules to determine a targeting strategy that obtains good Census coverage—and good characteristic enumeration—with the use of administrative records.

Highlights: Staff submitted to a peer-reviewed journal the accepted paper, "An Approach for Using Administrative Records to Reduce Contacts in the 2020 Decennial Census," that documents a scenario of administrative records vacancy determination and enumeration using a linear programming approach. Staff continued research on the comparison of classification methods for a person-place model for administrative records usage and submitted a paper on the topic to a peer-reviewed journal. Staff began work on developing extensions of the 2015 Census test methodology for the 2016 Census test.

Staff investigated decision theoretical strategies to exploit information on response propensity, along with information on administrative records (AR) quality, when conducting NRFU. The decision-theoretical strategies optimize field effort in concert with administrative record substitution to identify the most productive combination of fieldwork and AR utilization leading to the most accurate enumeration while containing costs. Staff presented early results at a Seminar in the series of study sessions on Bayesian decision theory and was invited to present this research at the gathering of international experts on Bayesian Adaptive Design Methods at the Census Bureau. Staff continues to explore strategies based on the propensity distributions observed in Census 2010 and in the tests leading to the 2020 Census. Staff is preparing a paper containing theoretical results in decision theory, as well as general operational recommendations for Census 2020.

Staff: Darcy Steeg Morris (x33989), Yves Thibaudeau

G. Evaluation of Response Error Using Administrative Records

Description: Censuses and their evaluations ask respondents to recall where they lived on Census Day, April 1. Some interviews for evaluations take place up to eleven months after this date. Respondents are asked when they moved to their current address, and the assumption has been that respondents who move around April 1 are able to give correct answers. Error in recalling a move or a move date may cause respondents to be enumerated at the wrong location in the census. This study investigates recall error in reports of moves and move dates in censuses and sample surveys using data from survey files linked to administrative records.

Highlights: Staff continues to collaborate with staff in the Census for Survey Measurement (CSM) on analyses of recall error for reports of moves and move dates in surveys using data from survey files linked to administrative records. Staff has pursued two studies: one uses data prepared for the "Memory Recall of

Migration Dates in the National Longitudinal Survey of Youth" developed under a contract with the National Opinion Research Center (NORC), and the other uses data from the Recall Bias Study, which was part of the 2010 Census Evaluation and Experiments Program. Staff continues to improve the drafts of two manuscripts, one for each study, through addressing the comments received from reviewers. In addition, staff presented an invited paper regarding lessons learned about evaluating survey data with administrative records files at the 2016 Methodology Symposium sponsored by Statistics Canada and is preparing a paper for the conference proceedings.

Staff: Mary Mulry (x31759)

H. Special Census: Disclosure Avoidance in Group Quarters

Description: Staff works with the Decennial Information Technology Division (DITD) to create synthetic data for disclosure avoidance in group quarters data for ongoing Special Census production.

Highlights: During Q1 and Q2 of FY 2016, staff continued to work with DITD in creating synthetic data for disclosure avoidance in group quarters data for ongoing Special Census production for certain localities in Iowa, Illinois, and Arizona. Staff determined which data were at potential risk of disclosure and applied statistical models to produce new data to replace those items. DITD integrated this data into the final product.

Staff: Rolando Rodriguez (x31816)

I. 2020 Unduplication Research

Description: The goal of this project is to conduct research to guide the development and assessment of methods for conducting nationwide matching and unduplication in the 2020 Decennial Census, future Censuses and other matching projects. Our staff will also develop and test new methodologies for unduplication. The project is coordinated by one of the 2020 Census Integrated Project Teams.

Highlights: No significant updates these quarters.

Staff: Michael Ikeda (x31756), Ned Porter, Bill Winkler, Emanuel Ben-David

1.3 ADDRESS CANVASSING IN FIELD (Decennial Project 6350B02)

A. Master Address File (MAF) Error Model and Quality Assessment

Description: The MAF is an inventory of addresses for all known living quarters in the U.S. and Puerto Rico. This project will develop a statistical model for MAF errors for housing units (HUs), group quarters (GQs), and transitory locations (TLs). This model, as well as an independent team, will be used to conduct independent quality checks on updates to the MAF and to ensure that these quality levels meet the 2020 Census requirements.

Highlights: Staff completed a research report on zero-inflated negative binomial modeling with exhaustive variable selection using 2010 address canvassing database plus several supplemental data sources. Staff switched to Bayesian modeling to facilitate more elaborate models and additional model outputs such as prediction intervals. Staff evaluated spatial count models to capture variability due to adjacency between blocks. Conditional autoregressive (CAR) models were considered. Staff also considered a methodology by Hughes and Haran to avoid confounding between spatial random effects and covariates, and to reduce the dimension of spatial effects. Initial spatial results show incremental improvements in predictions, but estimates of regression coefficients may change substantially from non-spatial models. Staff began to explore ordered categorical models as more robust alternatives to count models. Staff began to explore applications of statistical decision theory, and ways a model could assist in the latest operational plans.

Staff: Andrew Raim (x37894), Laura Ferreira (DSSD), Krista Heim (DSSD), Scott Holan (University of Missouri)

B. Development of Block Tracking Database

Description: The Targeted Address Canvassing (TRMAC) project supports Reengineered Address Canvassing for the 2020 Census. The primary goal of the TRMAC project is to identify geographic areas to be managed in the office (i.e., in-office canvassing) and geographic areas to be canvassed in the field. The focus of the effort is on decreasing in-field and assuring the Master Address File (MAF) is current, complete, and accurate. The Block Assessment, Research, and Classification Application (BARCA) is an interactive review tool which will allow analysts to assess tabulation blocks—and later Basic Collection Units (BCUs)—by comparing housing units in 2010 imagery and current imagery, along with TIGER reference layers and MAF data.

Highlights: During Q1 and Q2 of FY 2016, the BARCA in-office canvassing completed review of over 1.7 million tabulation blocks and loaded results into the Block Tracking Database (BTD). A quality control feature has been incorporated into BARCA that enables an adjudicator to assess any reviewer errors and correct them before results are loaded into the BTD. A system of reports has also been added to the system.

Staff: Tom Petkunas (x33216)

C. Detection of Map Changes
Description: This research is concerned with developing statistical techniques to detect changes in maps, utilizing remote sensing data, such as LIDAR.

Highlights: Staff (a) collaborated with Census geographers to develop methods for detecting map changes; (b) learned GIS (Geographic Information System) software to process data; and (c) improved road detection algorithm by adding a connectivity criterion to identify road networks.

Staff: Dan Weinberg (x38854)

1.4 DATA CODING, EDITING, AND IMPUTATION (Decennial Project 6550B01)

1.5 POLICY (Decennial Project 6250B07)

A. Privacy and Confidentiality for the 2020 Census
Description: This project undertakes research to understand privacy and confidentiality concerns related to methods of contact, response, and administrative records use which are under consideration for the 2020 Census. Methods of contact and response under consideration include internet alternatives such as social networking, email, and text messages. The project objectives are to determine privacy and confidentiality concerns related to these methods, and to identify a strategy to address the concerns.

Highlights: No significant updates these quarters.

Staff: Martin Klein (x37856)

1.6 AMERICAN COMMUNITY SURVEY (ACS) (Decennial Project 6385B70)

A. ACS Applications for Time Series Methods
Description: This project undertakes research and studies on applying time series methodology in support of the American Community Survey (ACS).

Highlights: Staff extended R code for custom multi-year estimates to handle large data frames and addressed input-output issues. Staff met with clients from the Veterans Administration to discuss practical aspects of the project such as how the format of the data is prepared for us and the desired output.

Staff: Tucker McElroy (x33227), Osbert Pang

B. ACS Imputation Research and Development
Description: The American Community Survey (ACS) process of editing and post-edit data-review is currently time and labor intensive. It involves repeatedly submitting an entire collection year of micro-data to an edit- enforcement program (SAS software). After each pass through the edit-enforcement program, a labor-intensive review process is conducted by a staff of analysts to identify inconsistencies and quality problems remaining in the micro-data. Before the data are ready for public release, they have at least three passes through the edit-enforcement program and three review processes by the analysts, taking upward of three months. The objective of this project is to experiment with a different strategy for editing—while keeping the same edit rules—and to assess if the new strategy can reduce the number of passes through the edit process and the duration of the review process.

Highlights: Staff completed a Research and Evaluation Analysis Plan (REAP) and obtained approval for the general research in software development. A new program of research is being developed.

Staff: Yves Thibaudeau (x31706), Rolando Rodriguez

C. Data Analysis of ACS CATI-CAPI Contact History
Description: This project continues earlier analyses of the American Community Survey (ACS) Computer Assisted Telephone Interview (CATI) and Computer Assisted Personal Interview (CAPI) contact history data. It focuses exclusively on CAPI with the goal of informing policy decisions on curtailing of CAPI contact attempts to minimize respondent burden on sampled households without unacceptable losses of ACS interviews.

Highlights: During Q1 and Q2 of FY 2016,, activity focused on analyzing the results of an ACS Pilot in which a new policy on centralized curtailment of CAPI case follow-up was tested in roughly one-quarter of CAPI cases in August 2015, in anticipation of national rollout of this policy. The goal of the policy was to reduce burden on potential respondents without major decreases in the CAPI case rate of interview completion. The policy had been developed based on earlier research in this project, and consisted of withdrawing a case from its field representative's (FR's) workload when a measure of the cumulative burden imposed on a potential respondent as a result of repeated contact attempts including attempts in pre-CAI modes) crossed a pre-set “maximum burden” threshold.

During this period, staff analyzed the pilot study results with the objective of comparing lost interviews and workload reduction with the levels forecast from analysis of previous (2012) ACS data, and of testing the comparative outcomes of workload and lowered interview completion rate between three “treatments”. The treatments consisted of implementation of the policy with and without telling the FRs each day what their current cumulative-burden score was, and of a control in which cases were not removed as a result of exceeding the cumulative-burden threshold. Results of the pilot were written into a completed and publicly released *ACS Research Report.*

Staff: Eric Slud (x34991), Robert Ashmead, Todd Hughes (ACSO), Rachael Walsh (OSCA)

D. Assessing Uncertainty in ACS Ranking Tables
Description: This project presents results from applying statistical methods which provide statements of how good the rankings are in the ACS Ranking Tables (see The Ranking Project: Methodology Development and Evaluation Research Section under Projects 0331000 and 0925000).

Highlights: [See General Research: Survey Sampling-Estimation and Modeling (C), The Ranking Project: Methodology Development and Evaluation]

Staff: Tommy Wright (x31702), Martin Klein, Brett Moran, Nathan Yau

E. Confidence Intervals for Proportions in ACS Data
[See General Research: Small Area Estimation (B), Coverage Properties of Confidence Intervals for Proportions in Complex Surveys]

F. Mode-Based Imputation in ACS
Description: It is well known that item nonresponse differs markedly between the different modes of data collection within the American Community Survey (ACS), yet current ACS practice is to perform item imputations via hot-deck methods pooling all ACS modes together. This project investigates the impact of doing ACS item-imputation separately within modes (Mail, CATI, CAPI) using the 2012 year of ACS data. It does this by developing a model-based mass-imputation approach to imputation using categorical control variables as similar as possible to those used in current hot-deck imputation and by comparing the impact on ACS estimates for selected survey variables if the mass imputation is done ignoring mode vs. being done within cells cross-classified by mode.

Highlights: This project is now complete.

Staff: Eric Slud (x34991), Ryan Janicki

G. Voting Rights Section 203 Model Evaluation and Enhancements Towards Mid-Decadal Determinations
Description: Section 203 of the *Voting Rights Act* asks for determinations relating to limited English proficiency and limited education of specified small domains (race and ethnicity groups) for small areas such as counties or minor civil divisions (MCDs). Section 203 seeks to determine whether or not small areas must provide voting materials in languages other than English. Previous research undertaken provided a small area model-based estimate derived from American Community Survey (ACS) 5-year data and 2010 Census data, which provides smaller estimated variances than ACS design-based estimates in many cases. Research and groundwork into the production mid-decade determination is ongoing.

Highlights: During Q1 and Q2 of FY 2016, staff evaluated aspects of the model applied during FY 2011 for the purpose of making determinations under Section 203. Evaluation itself has shifted to a prospective application of methods on ACS 5-year 2008-2012 data in lieu of the more recent 2010-2014 data slated for production. In this regard, two competing approaches to make estimations suitable for making determinations under Section 203 are approached based on different assumptions of what may or may not be important to the small area modeling question. Staff is currently working on both model evaluations and computational methods to give guidance to the appropriateness of particular models. Most is focused on comparing the strengths of competitor models as well as understanding their properties and failings. Some work has occurred on evaluating the utility and nature of design-effect adjustments with simulation study methods and with the co-involvement of methods proposed by Small Area Estimation Research Group staff. The current goal is a release of determinations by Q1 of FY 2017 with several implied milestones in between.

Staff: Patrick Joyce (x36793), Eric Slud, Robert Ashmead, Tommy Wright, Tom Louis (ADRM), John Abowd (ADRM)

1.7 DEMOGRAPHIC STATISTICAL METHODS DIVISION SPECIAL PROJECTS (Demographic Project TBA)

A. Special Project on Weighting and Estimation

Description: This project involves regular consulting with Current Population Survey (CPS) Branch staff on design, weighting, and estimation issues regarding the CPS. Issues discussed include design strategy for systematic sampling intervals, for rotating panels, composite estimation, variance estimation, and the possibility of altering CPS weighting procedures to allow for a single simultaneous stage of weight-adjustment for nonresponse and population controls.

Highlights: No significant updates these quarters.

Staff: Eric Slud (x34991), Yang Cheng (DSMD)

B. Weighted Estimating Equations with Response Propensities in Terms of Covariates Observed Only for Responders

Description: Regression survey estimators based on nonresponse adjustments are generally expressed through estimating equations in which survey weights are adjusted for nonresponse through estimated propensities. The propensities are generally functions of survey variables observed for all sampled individuals, yet the validity of the estimating equation estimators depends on an assumption that the outcomes and response indicators are conditionally independent, something that would be much more plausible if the propensities were allowed to depend also on variables involving demographics observed at the individual level only for survey-responders. This project is devoted to theoretical research on methods of weighted estimating equations applicable to surveys in which propensities may depend on covariates observed only for responders, which are illustrated on American Community Survey data.

Highlights: During Q1 and Q2 of FY 2016, staff continued research in this area, preparing an extensive data analysis using ACS 2012 data, to illustrate the potential benefits of the approach in reducing the bias of survey-estimated population totals for various attributes. Staff prepared and presented a talk on the theoretical underpinnings of this method at a Workshop on Nonignorable Nonresponse on November 12-13, 2015. Staff is continuing research in this area and preparing a manuscript for journal submission in the next quarter.

Staff: Eric Slud (x34991)

1.8 DEMOGRAPHIC SURVEYS DIVISION (DSD) SPECIAL PROJECTS (Demographic Project 0906/1444X00)

A. Data Integration

Description: The purpose of this research is to identify microdata records at risk of disclosure due to publicly available databases. Microdata from all Census Bureau sample surveys and censuses will be examined. Potentially linkable data files will be identified. Disclosure avoidance procedures will be developed and applied to protect any records at risk of disclosure.

Highlights: Staff worked with the Center for Disclosure Avoidance Research (CDAR) to continue planning stages to confirm suspected records for re-identification in the American Housing Survey. These stages include outlining software specifications for confirming suspicious records, identifying input file variables, and developing weighted likelihood measures of re-identification for the output file. Test runs of a beta algorithm on simulated data showed efficiency of the algorithm. Beta software is undergoing updates to improve efficiency. Additional software has been developed to clean the data.

Staff: Ned Porter (x31798), Marlow Lemons (CDAR)

1.9 POPULATION DIVISION PROJECTS (Demographic Project TBA)

A. Introductory Sampling Workshop

Description: In support of Population Division’s International Programs Area, staff will conduct (on request) introductory sampling workshops with focus on probability sampling for participants from various countries. These workshops are primarily funded by USAID.

Highlights: Over the two-week period (October 26-November 6, 2015), staff conducted a Workshop: Introduction to Survey Sampling (focus on Probability Sampling) at the Census Bureau Headquarters. The workshop presented the main components of survey sampling with a focus on probability sampling (and estimation) techniques. The hands-on, interactive workshop included the production of estimates of population parameters from sample surveys as a function of sample design, weighting procedures, the computation of sampling errors of sample estimators, and the making of inferences from the sample to the population. The seven workshop participants were mostly staff from statistical agencies in the United States, Ethiopia, Angola, and Namibia.

The workshop featured a Panel on Sampling (on the last day) which gave overviews of the American Community Survey, the Monthly/Annual Retail Trade Surveys, and the Current Population Survey.

Staff: Tommy Wright (x31702), Michael Leibert

1.10 SOCIAL, ECONOMIC, AND HOUSING STATISTICS DIVISION SMALL AREA ESTIMATION PROJECTS (Demographic Project 7165016)

A. Research for Small Area Income and Poverty Estimates (SAIPE)
Description: The purpose of this research is to develop, in collaboration with the Small Area Estimates Branch in the Social, Economic, and Housing Statistics Division (SEHSD), methods to produce “reliable” income and poverty estimates for small geographic areas and/or small demographic domains (e.g., poor children age 5-17 for counties). The methods should also produce realistic measures of the accuracy of the estimates (standard errors). The investigation will include assessment of the value of various auxiliary data (from administrative records or surveys) in producing the desired estimates. Also included would be an evaluation of the techniques developed, along with documentation of the methodology.

Highlights: During Q1 and Q2 of FY 2016, staff explored methodology for the estimation of year-to-year changes in rates of children in poverty through jointly modeling two years of ACS one-year estimates through a bivariate Fay-Herriot model with measurement error, where the measurement error variable comprises past estimates of children in poverty derived from ACS five-year data. The results were compared with results from a bivariate model that excluded the five-year estimates, and with a naïve model which treats the ACS five-year estimates as a covariate with no measurement error. These models were applied to the ACS estimates both with and without additional covariates from administrative records to understand the impact of these covariates. Staff implemented all of these models in a Bayesian setting initially using the generic software JAGS, using the prior distributions available there. Staff also developed a Markov Chain Monte Carlo algorithm tailored to this model, using a class of priors for which staff proved the propriety of the posterior under mild conditions.

Staff also finalized the paper on the Binomial Logit Normal (BLN) model applied to county rates of children in poverty, which is now published in a joint issue of *Statistics in Transition and Survey Methodology*. This work explores a model that attempts to address some problems with the current production model, such as the dropping of counties with zero observed children in poverty from the model fitting, the use of increasingly dated Census 2000 estimates as covariates, as well as the use of normality assumptions for inherently discrete data. It also compares the performance of bivariate and temporal extensions of the BLN model, when applied to children in poverty.

Staff is also exploring whether the BLN model may perform better than the current production model.

Staff: Jerry Maples (x32873), Carolina Franco, William Bell (ADRM)

B. Small Area Health Insurance Estimates (SAHIE)
Description: At the request of staff from the Social, Economic, and Housing Statistics Division (SEHSD), our staff will review current methodology for making small area estimates for health insurance coverage by state and poverty level. Staff will work on selected topics of SAHIE estimation methodology, in conjunction with SEHSD.

Highlights: During Q1 and Q2 of FY 2016, staff continued work on estimating year-to-year change of small area parameters in nonlinear small area models. The beta mixed effects regression model, with multivariate Gaussian errors for random effects from different time periods, was used for modeling direct estimates of the proportion with health insurance. The logit transformation of the direct estimates was used, so that the moments of the transformed variables could be expressed in terms of the digamma and trigamma functions. A literature review of inequalities and series expansions for the digamma and trigamma functions was done, and different moment estimators of the fixed effects and variance components based on these approximations were constructed. Simulation studies were conducted to investigate properties of these estimators.

Staff: Ryan Janicki (x35725)

C. Sub County Estimates of Poverty from Multi-year ACS Data
Description: This project is from the Development Case Proposal to improve the estimates of poverty related outcomes from the American Community Survey (ACS) at the tract level. Various modeling techniques, including model-based and model-assisted, will be used to improve on the design-based multi-year estimates currently produced by the ACS. The goal is to produce more accurate estimates of poverty and income at the tract level and develop a model framework that can be extended to outcomes beyond poverty and income.

Highlights: No significant updates these quarters.

Staff: Jerry Maples (x32873), Ryan Janicki, Carolina Franco, William Bell (ADRM)

1.11 ECONOMIC STATISTICAL COLLECTION (Economic Project 1183X01)

1.12 ECONOMIC MONTHLY/RETAIL (Economic Project 1001X00)

A. Research on Imputation Methodology for the Monthly Wholesale Trade Survey

Description: In the previous phase of this project, staff conducted a simulation study to investigate new imputation methodology for the Monthly Wholesale Trade Survey (MWTS). In this phase of the project, staff are creating a more realistic simulated wholesale trade population and investigating improved MWTS estimators. The MWTS is a longitudinal survey that provides month-to-month information on sales and inventories of U.S. merchant wholesalers. Key estimates produced from this survey include total sales, month-to-month relative change in sales, total inventories, and month-to-month relative change in inventories (overall and within industry subclasses). There are a number of challenges when developing estimators for the MWTS, including variables with highly skewed distributions, missing values in predictor variables from the Economic Census, and survey variables with trends that differ across industry classes. The longitudinal information in addition to a rich set of frame data available from the Economic Census can be used to build Bayesian models that address these challenges. It is expected that this model will be applicable to other business surveys.

Highlights: Staff are developing a new version of the realistic artificial population that is used to draw simulated samples of MWTS data. This version of the population will allow for distinctions between sampling units, reporting units, and tabulation units, thus allowing for more realistic simulations. Staff have also made some improvements to the roster of units in the population, as well as some improvements to the imputation methods used to fill in data for population units.

Staff: Martin Klein (x37856), Joe Schafer (ADRM), Joanna Fane Lineback (ADEP), Brett Moran

B. Use of Big Data for Retail Sales

Description: In this project, we are investigating the use of “Big Data” to fill gaps in retail sales estimates currently produced by the Census Bureau. Specifically, we are interested in how to use “Big Data” to supplement existing monthly/annual retail surveys with a primary focus on exploring (1) how to use third party data to produce geographic level estimates more frequently than once every five years (i.e. a new product), and (2) the possibility of using third party data tabulations to improve/enhance Census Bureau estimates of monthly retail sales - for example, validation and calibration. Various types of data are being pursued such as credit card transaction data and scanner data.

Highlights: During Q1 and Q2 of FY 2016, staff contributed to a final report and internal presentations of the evaluation of the quality of aggregated MasterCard data. Staff worked with Economic Directorate staff to assess the quality and potential use of aggregated electronic transaction data from First Data. Staff participated in working meetings with staff from Palantir, the company that houses, manages and visualizes the First Data data. Staff studied (1) small area estimation models for assessing the predictive power of the First Data transaction data for industry/state level estimates of monthly change and retail sales totals, and (2) the use of daily data to evaluate the Census Bureau's current trading day imputation procedure. Staff generated R routines to display high frequency time series and create moving holiday regressors that have similar properties to those of X-13ARIMA-SEATS. Staff adapted custom R software for time series modeling and signal extraction to study daily retail time series, successfully isolating trend, monthly seasonality, weekly seasonality, and an Easter effect, along with measures of uncertainty. It is conjectured that the weekly seasonality, once aggregated to a monthly level, corresponds to the trading day effect in monthly series.

Staff: Darcy Steeg Morris (x33989), Osbert Pang, Tommy Wright, Tucker McElroy, Brian Monsell, Bill Bostic (ADEP), Scott Scheleur (SSSD), Bill Davie, Jr. (SSSD)

1.13 ECONOMIC CENSUS/SURVEY ENGINEERING: TIME SERIES RESEARCH; ECONOMIC MISSING DATA/PRODUCT LINE DATA; DEVELOPMENT/SAS (Economic Project 2220B10)

A. Seasonal Adjustment Support

Description: This is an amalgamation of projects whose composition varies from year to year but always includes maintenance of the seasonal adjustment software used by the Economic Directorate.

Highlights: Staff provided support seasonal adjustment and software support for users within and outside the Census Bureau, including EJJE (Mexico), Epistemic, Ernst and Young, Goldman Sachs, M&T Bank, Obiettivo Lavoro (Italy), Palantir, Vanguard, Conference Board, New York Federal Researve Board, Bureau of Labor Statistics, European Central Bank, Swiss National Bank, KOF Swiss Economic Institute (Switzerland), National Bureau of Statistics (Nigeria), Office of National Statistics (UK), Statistics Canada, Statistics New Zealand, Columbia University, Northwestern, Catholic University of Louvain, and Zhejiang University.

Staff participated in a time series workshop in London, UK on November 19 and 20, 2015, sponsored by the Office of National Statistics. Staff presented the current state of time series research and development at the Census Bureau and participated in brainstorming sessions on future time series work.

Staff began working with Economic Directorate staff to plan a seasonal adjustment workshop to be held in November 2016.

Staff: Brian Monsell (x31721), James Livsey, Tucker McElroy, Osbert Pang, Anindya Roy, William R. Bell (ADRM)

B. Seasonal Adjustment Software Development and Evaluation

Description: The goal of this project is a multi-platform computer program for seasonal adjustment, trend estimation, and calendar effect estimation that goes beyond the adjustment capabilities of the Census X-11 and Statistics Canada X-11-ARIMA programs, and provides more effective diagnostics. The goals for FY 2015 include: continuing to develop a version of the X-13ARIMA-SEATS program with accessible output and updated source code so that, when appropriate, the Economic Directorate can produce SEATS adjustments; and incorporating further improvements to the X-13ARIMA-SEATS user interface, output and documentation. In coordination and collaboration with the Time Series and Related Methods Staff of the Economic Statistical Methods Division (ESMD), staff will provide internal and/or external training in the use of X-13ARIMA-SEATS and the associated programs, such as X-13-Graph, when appropriate.

Highlights: Staff released an updated version of X-13ARIMA-SEATS (Version 1.1, Build 26) to the Economic Directorate for testing. They compared adjustments from this version of the software to the last released version of X-13ARIMA-SEATS (Version 1.1, Build 19) and found, in all cases, no significant differences in the adjustments.

Staff implemented two new diagnostics for adequacy of residuals (Friedman's, Durbin Watson), added tables for outlier adjusted SEATS seasonal adjustment (irregular), added more diagnostics to .udg output, implemented the correct stock length of month regressors, and reformatted some of the HTML output.

After the Economic Directorate completed its testing, staff released Version 1.1, Build 26 of X-13ARIMA-SEATS to the public.

Staff is currently developing a library of the X-13ARIMA-SEATS routines, and adding an option to the spectrum spec to generate quarterly seasonality diagnostics for monthly series.

Staff continued the development of sigex, a suite of R routines for modeling multivariate time series. The software was extended to allow for estimation of fixed regression effects (which can be different for each series) and the software was applied to several daily time series.

Staff: Brian Monsell (x31721), Tucker McElroy, Osbert Pang

C. Research on Seasonal Time Series - Modeling and Adjustment Issues

Description: The main goal of this research is to discover new ways in which time series models can be used to improve seasonal and calendar effect adjustments. An important secondary goal is the development or improvement of modeling and adjustment diagnostics. This fiscal year’s projects include: (1) continuing research on goodness of fit diagnostics (including signal extraction diagnostics and Ljung-Box statistics) to better assess time series models used in seasonal adjustment; (2) studying the effects of model based seasonal adjustment filters; (3) studying multiple testing problems arising from applying several statistics at once; (4) determining if information from the direct seasonally adjusted series of a composite seasonal adjustment can be used to modify the components of an indirect seasonal adjustment, and more generally investigating the topics of benchmarking and reconciliation for multiple time series; (5) studying alternative models of seasonality, such as Bayesian and/or long memory models and/or heteroskedastic models, to determine if improvement to seasonal adjustment methodology can be obtained; (6) studying the modeling of stock holiday and trading day on Census Bureau time series; (7) studying methods of seasonal adjustment when the data are no longer univariate or discrete (e.g., multiple frequencies or multiple series); (8) studying alternative seasonal adjustment methods that may reduce revisions or have alternative properties; and (9) studying nonparametric methods for estimating regression effects, and their behavior under long range dependence and/or extreme values.

Highlights: Staff (a) conducted simulation and empirical work to vet new methodologies for fitting vector moving average models; (b) developed estimators for initial values needed to compute signal extraction estimates in a state space framework; (c) modeled weather data from the National Climatic Data Center (obtained through a web scraping tool) to be used in a weather-assisted seasonal adjustment of construction series; and (d) modeled daily time series (New Zealand immigration data, and credit card transaction data) with multiple forms of seasonality, and utilized software from the Bureau of Labor Statistics to obtain seasonal adjustments.

Staff: Tucker McElroy (x33227), James Livsey, Brian Monsell, Osbert Pang, William Bell (ADRM)

D. Supporting Documentation and Software for X-13ARIMA-SEATS

Description: The purpose of this project is to develop supplementary documentation and utilities for X-13ARIMA-SEATS that enable both inexperienced seasonal adjustors and experts to use the program as effectively as their backgrounds permit. This fiscal year's goals include improving the X-13ARIMA-SEATS documentation, exploring the use of R packages that interface with X-13ARIMA-SEATS, and exploring the use of component and Java software developed at the National Bank of Belgium.

Highlights: Staff updated the X-13ARIMA-SEATS REFERENCE MANUAL to include information on new options and diagnostics, and updated the "Getting Started" papers used to introduce users to X-13ARIMA-SEATS.

Staff updated HTML files to release new versions of X-13ARIMA-SEATS and Win X-13, and updated HTML documentation files for Win X-13.

Staff worked with Christoph Sax to improve utilities related to the seasonal R package. Staff also collaborated with Christoph Sax in creating an interface to improve the Census Bureau's communication of seasonal adjustment and X-13ARIMA-SEATS.

Staff updated software license and disclaimer statement after consulting with Commerce Department Lawyers.

Staff: Brian Monsell (x31721), Tucker McElroy, James Livsey, Osbert Pang, William R. Bell (ADRM), David Findley (Consultant)

E. Missing Data Adjustment Methods for Product Data in the Economic Census

Description: The Economic Census collects general items from business establishments such as total receipts, as well as more detailed items such as product sales. Although product data are an essential component of the Economic Census, item response rate is low. This project investigates methods for imputing missing product data in the Economic Census. Staff researched three methods for treating missing product line data: expansion estimation, hot deck (random and nearest neighbor), and sequential regression multivariate imputation (SRMI). Staff was asked to apply the SRMI method to these data and assist in making a recommendation.

Highlights: Staff was integrally involved in applying classification trees to determine characteristics of industries for which one variation of hot deck outperformed the other (random hot deck versus nearest neighbor hot deck). Staff worked with researchers from the Economic Directorate on a presentation and paper entitled, "Using Classification Trees to Recommend Hot Deck Imputation Methods: A Case Study" for the 2015 FCSM conference. Staff also presented this work internally to the editing and imputation knowledge sharing community group.

Staff: Darcy Steeg Morris (x33989), Maria Garcia, Yves Thibaudeau

1.14 2012 COMMODITY FLOW SURVEY (Economic Project 7103012)

A. 2012 Commodity Flow Survey

Description: This project provides a retrospective analysis of the cost-quality tradeoffs that the Commodity Flow Survey (CFS) made moving from a 2007 paper-only to a 2012 paper and electronic multi-mode data collection strategy. Based on the data quality findings, the possibility of adding additional edits or modifications to the instruments will be investigated. Optimization strategies for a multi-mode data collection strategy in the 2017 CFS and cost-quality implications of an all-electronic collection will be studied.

Highlights: Staff completed the final research report which summarized the results and recommendations from the project. This project is now complete.

Staff: Robert Ashmead (x31564), Eric Slud, Joanna Fane Lineback (ADEP)

1.15 INVESTIGATION OF ALTERNATIVE METHODS FOR RESOLVING BALANCE COMPLEX FAILURES IN StEPS (Economic Project TBA)

A. Investigation of Alternative Methods for Resolving Balance Complex Failures in StEPS

Description: The Standard Economic Processing System (StEPS) implements a raking algorithm for adjusting balance complexes in order to satisfy the requirement that the sum of items (details) in a balance complex balances to reported totals. In this project, we research alternative methods to raking when the data items are negative or when there is subtraction in the balance complex.

Highlights: One common method used in the StEPS generalized processing system is to rake details to the total for resolving failed balance complexes. This approach fails when data items can be negative or when there is subtraction in the balance complex as raking algorithms work for positive data only. This project has just started. Staff are in the process of understanding the problem, establishing project objectives, assigning tasks and researching alternative methods.

Staff: Maria Garcia (x31703), Yves Thibaudeau

1.16 BUSINESS DYNAMICS STATISTICS—EXPORT FILE WEIGHTING ISSUE (Research and Methodology Directorate TBA)

A. Business Dynamics Statistics—Export File Weighting Issue

Description: The challenge: we are unable to match the universe of export transactions to firms on the business register. Therefore, we cannot identify the universe of firms that export U.S. goods. We can pursue two options—(i) produce business dynamics statistics based on the identified cases only. For example, an official Census Bureau data product, the *Profile of U.S. Importing and Exporting Companies*, is released based on the "known" matches and users are provided with a technical documentation explaining the data limitations; or (ii) construct weights to create business dynamics statistics that are representative of the U.S. exporter population.

Highlights: During Q1 and Q2 of FY 2016, staff met bi-weekly with Center for Economic Studies staff.

Staff: Maria Garcia (x31703), Emanuel Ben-David

1.17 NATIONAL SURVEY OF DRUG USE & HEALTH (Census Bureau Project 7236045)

A. National Survey of Drug Use & Health

Description: This project is a feasibility study concerning the extension of the National Survey of Drug Use & Health (NSDUH) to Puerto Rico and other U.S. island areas. Our staff will focus specifically on small area estimation methodology and will determine if and how the island areas can be incorporated into the current NSDUH small area estimation methodology.

Highlights: Staff completed and submitted a draft of a report for the Substance Abuse and Mental Health Services Administration (SAMHSA) detailing the considerations for conducting the National Survey on Drug Use and Health in the U.S. island areas under consideration.

Staff: Robert Ashmead (x31564), Jerry Maples

1.18 PROGRAM DIVISION OVERHEAD (Census Bureau Project 0331000)

A. Center Leadership and Support

This staff provides ongoing leadership and support for the overall collaborative consulting, research, and administrative operation of the center.

Staff: Tommy Wright (x31702), Alisha Armas, Lauren Emanuel, Michael Hawkins, Michael Leibert, Erica Magruder, Eric Slud, Kelly Taylor

B. Research Computing

Description: This ongoing project is devoted to ensuring that Census Bureau researchers have the computers and software tools they need to develop new statistical methods and analyze Census Bureau data.

Highlights: During Q1 and Q2, the Integrated Research Environment (IRE) team continued to develop the IRE, a shared Linux computing platform that will replace the current "compute clusters" research1, research2, and the RDC cluster. The IRE will provide the logical separation of project data and activities that is currently provided in the RDC environment, but without using a separate login for each project. A collection of scripts will enable the user to "change into" a particular project where they will be presented only with the data associated with that project. Testing of those scripts and integrating them with the job scheduler (PBSPro) is the current focus. The current model for system use is for each user to have a separate window containing a Linux desktop session corresponding to that project. If working on more than one project at a time, the user will have multiple desktop windows open that they can switch between, but they will not be able to copy data between windows.

Once the basic environment is fully tested, focus will shift toward integrating the IRE with the Data Management System (DMS). IRE is integrated with the Center for Economic Studies (CES) management system (CMS). All access to data within the IRE will be specified by the CMS and DMS. When a project manager wants to add a person to a project, they will do so in CMS or DMS, and the IRE will reflect the changes. Initial migration of the RDC environment to the IRE is expected by the end of 2016, followed by the internal clusters (research1 and research2).

Staff: Chad Russell (x33215)

2. RESEARCH

2.1 GENERAL RESEARCH AND SUPPORT (Census Bureau Project 0331000)

2.2 GENERAL RESEARCH (Census Bureau Project 0925000)

Missing Data, Edit, and Imputation

Motivation: Missing data problems are endemic to the conduct of statistical experiments and data collection projects. The instigators almost never observe all the outcomes they had set to record. When dealing with sample surveys or censuses that means individuals or entities in the survey omit to respond, or give only part of the information they are being asked to provide. In addition the information provided may be logically inconsistent, which is tantamount to missing. To compute official statistics, agencies need to compensate for missing data. Available techniques for compensation include cell adjustments, imputation and editing. All these techniques involve mathematical modeling along with subject matter experience.

Research Problems: Compensating for missing data typically involves explicit or implicit modeling. Explicit methods include Bayesian multiple imputation and propensity score matching. Implicit methods revolve around donor-based techniques such as hot-deck imputation and predictive mean matching. All these techniques are subject to edit rules to ensure the logical consistency of remedial product. Research on integrating together statistical validity and logical requirements into the process of imputing continues to be challenging. Another important problem is that of correctly quantifying the reliability of predictors that have been produced in part through imputation, as their variance can be substantially greater than that computed nominally.

Potential Applications: Research on missing data leads to improved overall data quality and predictors accuracy for any census or sample survey with a substantial frequency of missing data. It also leads to methods to adjust the variance to reflect the additional uncertainty created by the missing data. Given the ever rising cost of conducting censuses and sample surveys, imputation and other missing-data compensation methods may come to replace actual data collection, in the future, in situations where collection is prohibitively expensive.

A. Editing
Description: This project covers development of methods for statistical data editing. Good methods allow us to produce efficient and accurate estimates and higher quality microdata for analyses.

Highlights: No significant updates these quarters.

Staff: Maria Garcia (x31703)

B. Editing and Imputation
Description: Under this project, our staff provides advice, develops computer edit/imputation systems in support of demographic and economic projects, implements prototype production systems, and investigates edit/imputation methods.

Highlights: Staff revised a paper submitted for publication. The paper discusses modeling conditional probability to predict missing variables in subsequent waves of a longitudinal survey.

Staff: Yves Thibaudeau (x31706), Maria Garcia, Martin Klein, Darcy Steeg Morris, Bill Winkler

Record Linkage

Motivation: Record linkage is intrinsic to efficient, modern survey operations. It is used for unduplicating and updating name and address lists. It is used for applications such as matching and inserting addresses for geocoding, coverage measurement, Primary Selection Algorithm during decennial processing, Business Register unduplication and updating, re-identification experiments verifying the confidentiality of public-use microdata files, and new applications with groups of administrative lists. Significant theoretical and algorithmic progress (Winkler 2004ab, 2006ab, 2008, 2009a; Yancey 2005, 2006, 2007, 2011) demonstrates the potential for this research. For cleaning up administrative records files that need to be linked, theoretical and extreme computational results (Winkler 2010, 2011b) yield methods for editing, missing data and even producing synthetic data with valid analytic properties and reduced/eliminated re-identification risk. Easy means of constructing synthetic make it straightforward to pass files among groups.

Research Problems: The research problems are in three major categories. First, we need to develop effective ways of further automating our major record linkage operations. The software needs improvements for matching large sets of files with hundreds of millions of records against other large sets of files. Second, a key open research question is how to effectively and automatically estimate matching error rates. Third, we need to investigate how to develop effective statistical analysis tools for analyzing data from groups of administrative records when unique identifiers are not available. These methods need to show how to do correct demographic, economic, and statistical analyses in the presence of matching error.

Potential Applications: Presently, the Census Bureau is contemplating or working on many projects involving record linkage. The projects encompass the Demographic, Economic, and Decennial areas.

A. Disclosure Avoidance for Microdata
Description: Our staff investigates methods of microdata masking that preserves analytic properties of public-use microdata and avoid disclosure.

Highlights: Staff reviewed twelve papers on variants of differential privacy and their relationship to Yang, Fienberg, and Rinaldo (2012) and Winkler (2010). Staff e-mailed comments to staff in the Center for Disclosure Avoidance Research (CDAR) on how the methods of modeling/edit/imputation in Winkler (1997, 2003, 2008, 2010) can be used for generating synthetic data with valid analytic properties and very significantly reduced re-identification risk. Staff also e-mailed several papers and an extensive lists of references on microdata confidentiality to staff in CDAR. Staff e-mailed information and a paper on microdata confidentiality to staff in the Economic Directorate.

Staff: William Winkler (x34729)

B. Record Linkage and Analytic Uses of Administrative Lists
Description: Under this project, staff will provide advice, develop computer matching systems, and develop and perform analytic methods for adjusting statistical analyses for computer matching error.

Highlights: Staff reviewed papers by Hof and Zwinderman (2012, 2015) and Tancredi and Liseo (2015) that had models for adjusting statistical analyses for linkage error. Staff reviewed and sent comments to the authors of three record linkage papers that had already appeared in journals.

Staff e-mailed comments to staff at the National Agriculture Statistics Service who are building a system for the 2017 Agriculture Census. Staff provided advice and an extensive lists of references to staff in the Center for Administrative Records Research and Applications (CARRA) for a proposed record linkage application. Because we are not SAS programmers, we were unable to write SAS software for the application. Staff spent four days finding record linkage software and non-Title 13 files for a demonstration in a Data Integration class at the University of Maryland, College Park (UMD) after backups on Census Bureau computers were lost after new PC backup software failed. Staff e-mailed the current versions of *BigMatch* that were used for 2010 Decennial Census production to staff in the Decennial Statistical Studies Division (DSSD). Staff also gave a large number of comments regarding machine learning and distributed computing research to staff in DSSD.

Staff circulated a review to staff in CARRA, DSSD, DITD, and CSRM indicating that *BigMatch* continues as the fastest record linkage software in the world (50 times as fast as parallel software from CS researchers at Stanford) and is generally the most accurate record linkage software (according to an extensive review of record linkage shareware and commercial software from SAS and IBM by Professor Anna Ferrante for a consortium of health agencies). Staff provided very extensive advice to a researcher at the National Institutes of Health on record linkage software and methods. Staff e-mailed information and comments related to cleaning up and analyzing national files to staff in the Computer Services Division (CSVD). Staff e-mailed a number of comments on edit/imputation systems to individuals in the CEDCAP project on edit/imputation. Staff did extensive reading of background literature on record linkage.

Staff: William Winkler (x34729), Ned Porter

C. Modeling, Analysis, and Quality of Data
Description: Our staff investigates methods of the quality of microdata primarily via modeling methods and new software techniques that accurately describe one or two of the analytic properties of the microdata.

Highlights: Staff e-mailed extensive comments to a professor at the University of Maryland, College Park (UMD) related to how the record linkage methods in Winkler (1994, 2008) were developed. Unlike conventional statistics, we develop an operational generalized system and then copy the theoretical likelihood development from the computer code. Staff answered many questions from students in UMD's Data Integration Topics class and gave demonstrations of two of the software packages (Yancey and Winkler 2004, 2009) and Winkler (2008, 2010).

Staff provided comments, advice, a list of references, and a list of shareware/freeware to the New York City government. Staff provided comments to staff at the Office of National Statistics in the UK related to twelve questions they had on record linkage. One staff member agreed to return to the Isaac Newton Institute at Cambridge University to give a talk on record linkage and participate in some research. Staff e-mailed comments to a professor and graduate student at Carnegie-Mellon University (CMU). Staff made comments to DSSD staff related to a research proposal from CMU. Staff provided extensive comments to staff of the Office of National Statistics in the UK related to the EM algorithm for parameter estimation in record linkage.

Staff completed the first software version of set covering algorithms edit generation for a program. The theory is based on Fellegi and Holt (*JASA* 1976), Garfinkel, Kunnathur, and Liepins (*Operations Research* 1986), and Winkler (1997). The first version works on the example of Garfinkel et al. Staff are working on a larger example for the Italian Labour Force Survey using data provided by IBM and ISTAT.

Staff worked on the following problem posed by a colleague: Given two real sequences $y_1 < y_2 < \ldots < y_N$ and $z_1 < z_2 < \ldots < z_n$, where $n < N$. We want to find a sequence $x_1 < x_2 < \ldots < x_n$ to minimize the value: $(z_1 - x_1)^2 + \ldots + (z_n - x_n)^2$ where x_i is in $\{y_1, \ldots, y_N\}$. Staff provided a heuristic solution using divide and conquer strategy. Based on a recommendation from another colleague, staff read a few chapters of *Numerical Optimization* about quadratic programming.

Staff provided extensive background on modeling/edit/imputation to individuals working on the CEDCAP edit/imputation project to develop generalized methods/software for the Decennial Census and, possibly, approximately 140 other Census Bureau surveys. The background document included a description of the specific work successfully performed at five statistical agencies that have been able to develop Fellegi-Holt systems. The background covered some of the specifics of the computational algorithms and gave a number of references in refereed journals and in agency technical reports. Staff also provided a document on how to develop teams with the technical skills for generalized systems based on successful projects at Statistics Canada and the Census Bureau: Winkler, W. E. and Hidiroglou, M. (1998), "Developing Analytic Programming Ability to Empower the Survey Organization," http://www.census.gov/srd/papers/pdf/rr9804.pdf .

Staff: William Winkler (x34729), Ned Porter, Maria Garcia

D. R Users Group
Description: The initial objective of the R Users Group is to identify the areas of the Census Bureau where R software is developed and those other areas that could benefit from such development. The scope of the topics is broad and it includes estimation, missing data methods, statistical modeling, Monte-Carlo and resampling methods. The ultimate goal is to move toward integrated R tools for statistical functionality at the Census Bureau.

Initially the group will review basic skills in R and provide remedial instruction as needed. The first topic for deeper investigation is complex-survey infrastructure utilities, in particular an evaluation of the "Survey" package and its relevance at the Census Bureau in the context of weighing, replication, variance estimation and other structural issues.

Highlights: No activity these quarters.

Staff: Yves Thibaudeau (x31706), Chad Russell

Small Area Estimation

Motivation: Small area estimation is important in light of a continual demand by data users for finer geographic detail of published statistics. Traditional demographic surveys designed for national estimates do not provide large enough samples to produce reliable direct estimates for small areas such as counties and even most states. The use of valid statistical models can provide small area estimates with greater precision, however bias due to an incorrect model or failure to account for informative sampling can result. Methods will be investigated to provide estimates for geographic areas or subpopulations when sample sizes from these domains are inadequate.

Research Problems:

- Development/evaluation of multilevel random effects models for capture/recapture models.
- Development of small area models to assess bias in synthetic estimates.
- Development of expertise using nonparametric modeling methods as an adjunct to small area estimation models.
- Development/evaluation of Bayesian methods to combine multiple models.
- Development of models to improve design-based sampling variance estimates.
- Extension of current univariate small-area models to handle multivariate outcomes.

Potential Applications:

- Development/evaluation of binary, random effects models for small area estimation, in the presence of informative sampling, cuts across many small area issues at the Census Bureau.
- Using nonparametric techniques may help determine fixed effects and ascertain distributional form for random effects.
- Improving the estimated design-based sampling variance estimates leads to better small area models which assumes these sampling error variances are known.
- For practical reasons, separate models are often developed for counties, states, etc. There is a need to coordinate the resulting estimates so smaller levels sum up to larger ones in a way that correctly accounts for accuracy.
- Extension of small area models to estimators of

design- base variance.

A. Small Area Methods with Misspecification
Description: In this project, we undertake research on area-level methods with misspecified models, primarily directed at development of diagnostics for misspecification using robust sandwich-formula variances, cross-validation, and others, and on Bayesian estimation of model parameters within two-component Fay-Herriot models.

Highlights: During Q1 and Q2 of FY 2016, no progress was made due to departure of key staff in the Center for Disclosure Avoidance Research Division (CDAR). Efforts are currently underway to obtain proper staffing on this project.

Staff: Jerry Maples (x32873), Gauri Datta, Eric Slud, Jiashen You (CDAR)

B. Coverage Properties of Confidence Intervals for Proportions in Complex Surveys
Description: This is primarily a simulation project to investigate the coverage behavior of confidence intervals for proportions estimated in complex surveys. The goal is ultimately to inform recommendations for interval estimates in the American Community Survey (ACS), so the issues of main interest are:
(i) whether the current Wald-type intervals (defined as a point-estimator plus or minus a margin-or-error (MOE) estimate) can be improved by empirical-Bayes modifications or by modified forms of intervals known to perform well in the setting of binomial proportion-estimators, (ii) whether failures of coverage in a simulated complex survey can be ascribed to poor estimation of effective sample size or to other aspects of inhomogeneity and clustering in proportions within realistically complex populations, and (iii) whether particular problems arising with coverage of intervals for small proportions can be overcome. Future research might address whether the confidence interval methods developed for single-domain design-based estimates can also be adapted to small area estimates that borrow strength across domains.

Highlights: Staff expanded the simulation study to include scenarios with more clustering and a higher intra-cluster correlation (ICC) in response to a recent related paper by Dean and Pagano (2015). Staff also incorporated two modifications to the effective sample size available in the literature, one proposed by Korn and Graubard (1998) and another by Dean and Pagano (2015). Staff began analysis of results under this more comprehensive simulation design. Staff studied additional questions such as the effect of the ICC on the coverage and length of all intervals, the effect of having larger cluster sizes, and the interaction of both effects. Staff derived two new methods of computing the design effect, which will be empirically evaluated through the simulation study.

Staff: Carolina Franco (x39959), Eric Slud, Thomas Louis (ADRM), Rod Little (University of Michigan)

C. Small Area Estimates of Disability
Description: This project is from the Development Case proposal to create subnational estimates of specific disability characteristics (e.g., number of people with autism). This detailed data is collected in a supplement of the Survey of Income and Program Participation (SIPP). However, the SIPP is only designed for national level estimates. This project is to explore small area models to combine SIPP with the large sample size of the American Community Survey to produce state and county level estimates of reasonable quality.

Highlights: During Q1 and Q2 of FY 2016, staff rewrote the codebase to have more flexibility in modelling options. Staff worked on and submitted a manuscript to the Small Area Special Edition of the *Journal of the Royal Statistical Society, Series A.*

Staff: Jerry Maples (x32873), Amy Steinweg (SEHSD)

D. Using ACS Estimates to Improve Estimates from Smaller Surveys via Bivariate Small Area Estimation Models
Description: Staff will investigate the use of bivariate area-level models to improve small area estimates from one survey by borrowing strength from related estimates from a larger survey. In particular, staff will explore the potential of borrowing strength from estimates from the American Community Survey, the largest U.S. household survey, to improve estimates from smaller U.S. surveys, such as the National Health Interview Survey, the Survey of Income and Program Participation, and the Current Population Survey.

Highlights: Staff prepared and delivered a presentation on related research results at the 2016 Ross-Royall Symposium at Johns Hopkins University.

Staff: Carolina Franco (x39959), William R. Bell (ADRM)

E. Multivariate Fay-Harriot Hierarchical Bayesian Estimation of Small Area Means under Functional Measurement Error
Description: Area-level models have been extensively used in small area estimation to produce model-based estimates of a population characteristic for small areas (e.g., Fay and Herriot, 1979). Multivariate area level models have also been used to jointly model multiple characteristics of correlated responses (e.g., Huang and Bell, 2012, Franco and Bell, 2015). Such models may lead to more precise small area estimates than separate univariate modeling of each characteristic. Typically both univariate and multivariate small area estimation

models use auxiliary information to borrow strength from other areas and covariates associated with a response variable or a response vector. However, auxiliary variables are sometimes measured or obtained from sample surveys and are subject to measurement or sampling error. Researchers recognized that ignoring measurement error in the covariates and using standard solutions developed for covariates measured without error may lead to suboptimal inference. It was demonstrated in the univariate small area estimation setup that this naïve approach can result in model-based small area estimators that are more variable than the direct estimators when some of the covariate values in a small area are measured with substantial error (cf. Ybarra and Lohr, 2008, *Biometrika*; Arima, Datta and Liseo, 2015, *Scandinavian Journal of Statistics*). We are investigating a multivariate Fay-Herriot model and develop Bayes small area estimates when one or more auxiliary variables are measured with error. We work out a hierarchical Bayesian analysis for the multivariate Fay-Herriot model with a functional measurement error treatment for the covariates measured with error.

Highlights: During Q1 and Q2 of FY 2016, staff investigated the performance of the multivariate Fay-Herriot measurement error model through applying it to estimating children in poverty in counties and to estimating median incomes for families using the software JAGS. Staff empirically compared a bivariate measurement error model to a naïve model where the covariate measured with error is treated as a known covariate. It is unclear whether it is possible to use JAGS to implement this model for problems with higher dimensions, and it is also not straightforward to apply a class of priors that staff proved to be proper under mild conditions. Staff developed a Markov Chain Monte Carlo algorithm customized to this problem using Metropolis Hastings steps and Gibbs sampling. Staff began programming the algorithm and debugging it. Staff also worked on a draft of a paper to be submitted to a journal, which is currently in progress. Staff presented preliminary results in an invited talk in the 8th International Conference of the ERCIM Working Group on Computational and Methodological Statists (CMStatistics 2015) at the University of London, United Kingdom.

Staff: Carolina Franco (x39959), Gauri Datta, William R. Bell (ADRM)

F. **Smoothing Design Effects for Small Sample Areas**

Description: In Small Area Estimation, the design-based estimates for many areas are based on very small samples. We propose using information from a larger aggregate, whose design-based variance estimator can be reliably estimated to inform us about the design effect at the small component area. Our goal is to create a principled method to use information about design effects at the higher level to estimate design effects at the lower level. Due to the lack of data, this will require strong assumptions and large amounts of smoothing of design features over the small local areas.

Highlights: During Q1 and Q2 of FY 2016, staff created a framework based on a pseudo stratified design to link the design effects at the local area to the variance estimate of the larger aggregate area. In order to preserve as much design information about the local area as possible, effects for unequal sample size, clustering structure and informative sampling (survey outcome related to probability of selection) are first conditioned out so that the residual design effect is what will be estimated from the higher level aggregate. Applications for this method are being investigated for the Voting Rights Act, SIPP state level variances and SAIPE sub-county level variances.

Staff: Jerry Maples (x32873)

Survey Sampling-Estimation and Modeling

Motivation: The demographic sample surveys of the Census Bureau cover a wide range of topics but use similar statistical methods to calculate estimation weights. It is desirable to carry out a continuing program of research to improve the accuracy and efficiency of the estimates of characteristics of persons and households. Among the methods of interest are sample designs, adjustments for non-response, proper use of population estimates as weighting controls, small area estimation, and the effects of imputation on variances.

The Economic Directorate of the Census Bureau encounters a number of issues in sampling and estimation in which changes might increase the accuracy or efficiency of the survey estimates. These include, but are not restricted to, a) estimates of low-valued exports and imports not currently reported, b) influential values in retail trade survey, and c) surveys of government employment.

The Decennial Census is such a massive undertaking that careful planning requires testing proposed methodologies to achieve the best practical design possible. Also, the U.S. Census occurs only every ten years and is the optimal opportunity to conduct evaluations and experiments with methodologies that might improve the next census. Sampling and estimation are necessary components of the census testing, evaluations, and experiments. The scale and variety of census operations require an ongoing research program to achieve improvements in methodologies. Among the methods of interest are coverage measurement sampling and estimation, coverage measurement evaluation, evaluation of census operations, uses of administrative records in census operations, improvements in census processing, and analyses that aid in increasing census response.

Research Problems:

- How can methods making additional use of administrative records, such as model-assisted and balanced sampling, be used to increase the efficiency of household surveys?
- Can non-traditional design methods such as adaptive sampling be used to improve estimation for rare characteristics and populations?
- How can time series and spatial methods be used to improve ACS estimates or explain patterns in the data?
- Can generalized weighting methods be implemented via optimization procedures that allow better understanding of how the various steps relate to each other?
- Some unusual outlying responses in the surveys of retail trade and government employment are confirmed to be accurate, but can have an undesired large effect on the estimates - especially estimates of change. Procedures for detecting and addressing these influential values are being extended and examined through simulation to measure their effect on the estimates, and to determine how any such adjustment best conforms with the overall system of estimation (monthly and annual) and benchmarking.
- What models aid in assessing the combined effect of all the sources of estimable sampling and nonsampling error on the estimates of population size?
- How can administrative records improve census coverage measurement, and how can census coverage measurement data improve applications of administrative records?
- What analyses will inform the development of census communications to encourage census response?
- How should a national computer matching system for the Decennial Census be designed in order to find the best balance between the conflicting goals of maximizing the detection of true duplicates and minimizing coincidental matches? How does the balance between these goals shift when modifying the system for use in other applications?
- What can we say about the additional information that could have been obtained if deleted census persons and housing units had been part of the Census Coverage Measurement (CCM) Survey?

Potential Applications:

- Improve estimates and reduce costs for household surveys via the introduction of additional design and estimation procedures.
- Produce improved ACS small area estimates through the use of time series and spatial methods.
- Apply the same weighting software to various surveys.
- New procedures for identifying and addressing influential values in the monthly trade surveys could provide statistical support for making changes to weights or reported values that produce more accurate estimates of month-to-month change and monthly level. The same is true for influential values in surveys of government employment.
- Provide a synthesis of the effect of nonsampling errors on estimates of net census coverage error, erroneous enumerations, and omissions and identify the types of nonsampling errors that have the greatest effects.
- Describe the uncertainty in estimates of foreign-born immigration based on American Community Survey (ACS) used by Demographic Analysis (DA) and the Postcensal Estimates Program (PEP) to form estimates of population size.
- Improve the estimates of census coverage error.
- Improve the mail response rate in censuses and thereby reduce the cost.
- Help reduce census errors by aiding in the detection and removal of census duplicates.
- Provide information useful for the evaluation of census quality.
- Provide a computer matching system that can be used with appropriate modifications for both the Decennial Census and several Decennial-related evaluations.

A. Household Survey Design and Estimation
[See Demographic Projects]

B. Sampling and Estimation Methodology: Economic Surveys

Description: The Economic Directorate of the Census Bureau encounters a number of issues in sampling and estimation in which changes might increase the accuracy or efficiency of the survey estimates. These include estimates of low-valued exports not currently reported, alternative estimation for the *Quarterly Financial Report*, and procedures to address nonresponse and reduce respondent burden in the surveys. Further, general simulation software might be created and structured to eliminate various individual research efforts. An observation is considered influential if the estimate of total monthly revenue is dominated by its weighted contribution. The goal of the research is to find methodology that uses the observation but in a manner that assures its contribution does not dominate the estimated total or the estimates of period-to-period change.

Highlights: Staff continued collaborating with a team in the Economic Directorate to find a statistical procedure for detecting and treating verified influential values in economic surveys to replace the current subjective procedure performed by analysts. Recent research has focused on finding an automated procedure with the expectation that any adjustments be reviewed. Previous research identified an M-estimation methodology as the most suitable choice, but the initial parameter settings for the M-estimation algorithm affect its performance. Using historical data from the Monthly Wholesale

Trade Survey (MWTS), staff developed an automated data-driven approach for determining the initial parameter settings for the M-estimation algorithm parameters. The next step is to use the method in a side-by-side test conducted in real time during MWTS data collection.

In addition, the team completed a second revision of a research note on the Clark method of Winsorization, an alternative method for detecting and treating influential values that the team investigated before deciding to pursue the M-estimation method. In the research note, the team presents the insights gained about the performance of Clark Winsorization in detecting influential values.

Staff: Mary Mulry (x31759)

C. The Ranking Project: Methodology Development and Evaluation

Description: This project undertakes research into the development and evaluation of statistical procedures for using sample survey data to rank several populations with respect to a characteristic of interest. The research includes an investigation of methods for quantifying and presenting the uncertainty in an estimated ranking of populations. As an example, a series of ranking tables are released from the American Community Survey in which the fifty states and the District of Columbia are ordered based on estimates of certain characteristics of interest.

Highlights: Staff worked on drafts of three visualizations of rankings using 2011 American Community Survey (ACS) "Travel Time to Work" data. One of the visualizations provides comparisons of pairs of means for the 51 (including Washington, D.C.) states; the second shows a bootstrap distribution for the estimated rank of each state; and the third visualization shows the bootstrap estimates of probability that the estimated rank for state *i* exceeds the estimated rank for state *j*. Preparation for an internet website began.

Staff: Tommy Wright (x31702), Martin Klein, Jerzy Wieczorek (Carnegie Mellon University), Brett Moran, Nathan Yau, Michael Leibert

D. Sampling and Apportionment

Description: This short-term effort demonstrated the equivalence of two well-known problems–the optimal allocation of the fixed overall sample size among L strata under stratified random sampling and the optimal allocation of the H = 435 seats among the 50 states for the apportionment of the U.S. House of Representatives following each decennial census. This project continues development with new sample allocation algorithms.

Highlights: Staff drafted a paper detailing exact optimal allocation algorithms given stated desired precision and given fixed budget. This project continues development with new allocation algorithms.

Staff: Tommy Wright (x31702), Andrew Perry

E. Interviewer-Respondent Interactions: Gaining Cooperation

Description: Survey nonresponse rates have been increasing, leading to concerns about the accuracy of (demographic) sample survey estimates. For example, from 1990 to 2004, initial contact nonresponse rates have approximately doubled for selected household sample surveys including the Current Population Survey (CPS) (from 5.7 percent to 10.1 percent). While mailout/mailback is a relatively inexpensive data collection methodology, decreases in mailback rates to censuses and sample surveys mean increased use of methodologies that bring respondents into direct contact with Census Bureau interviewers (e.g., field representatives) using CATI (computer assisted telephone interviewing) or CAPI (computer assisted personal interviewing). CAPI can include face-to-face or telephone contact. Unsuccessful interviewer-respondent interactions can lead to increased costs due to the need for additional follow-up, and can also decrease data quality. So, unsuccessful interviewer-respondent interactions should be minimized.

This project will analyze data from 512 field representatives (interviewers) as part of an exploratory study, examining their beliefs regarding what works in gaining respondents' cooperation and investigating associations with field representatives' performance in terms of completed interview rates. We will also study associations between field representatives' beliefs and what they say they do.

Highlights: No significant updates these quarters.

Staff: Tommy Wright (x31702), Tom Petkunas

Time Series and Seasonal Adjustment

Motivation: Seasonal adjustment is vital to the effective presentation of data collected from monthly and quarterly economic surveys by the Census Bureau and by other statistical agencies around the world. As the developer of the X-13ARIMA-SEATS Seasonal Adjustment Program, which has become a world standard, it is important for the Census Bureau to maintain an ongoing program of research related to seasonal adjustment methods and diagnostics, in order to keep X-13ARIMA-SEATS up-to-date and to improve how seasonal adjustment is done at the Census Bureau.

Research Problems:

- All contemporary seasonal adjustment programs of interest depend heavily on time series models for trading day and calendar effect estimation, for modeling abrupt changes in the trend, for providing required forecasts, and, in some cases, for the seasonal adjustment calculations. Better methods are needed for automatic model selection, for detection of inadequate models, and for assessing the uncertainty in modeling results due to model selection, outlier identification and non-normality. Also, new models are needed for complex holiday and calendar effects.
- Better diagnostics and measures of estimation and adjustment quality are needed, especially for model-based seasonal adjustment.
- For the seasonal, trading day and holiday adjustment of short time series, meaning series of length five years or less, more research into the properties of methods usually used for longer series, and perhaps into new methods, are needed.

Potential Applications:

- To the effective presentation of data collected from monthly and quarterly economic surveys by the Census Bureau and by other statistical agencies around the world.

A. Seasonal Adjustment
Description: This research is concerned with improvements to the general understanding of seasonal adjustment and signal extraction, with the goal of maintaining, expanding, and nurturing expertise in this topic at the Census Bureau.

Highlights: Staff (a) completed a simulation study for extreme value adjustment of New Zealand agricultural time series; (b) implemented seasonal heteroscedasticity models to show improved forecasting and seasonal adjustment of construction series; and (c) extended work on signal extraction decompositions allowing for correlation between components.

Staff: Tucker McElroy (x33227), James Livsey, Brian Monsell, Osbert Pang, Anindya Roy

B. Time Series Analysis
Description: This research is concerned with broad contributions to the theory and understanding of discrete and continuous time series, for univariate or multivariate time series. The goal is to maintain and expand expertise in this topic at the Census Bureau.

Highlights: Staff (a) extended work on stable parametrizations of VARMA models fitted under parameter constraints, utilizing a LASSO objective function; (b) continued simulation and software development for multivariate count time series; (c) further developed likelihood ratio tests for Granger non-causality, as a way to exclude extraneous data from multivariate forecasting problems; (d) continued research and simulations for two tests of co-integration, one based upon fitted structural models and another based on nonparametric spectral estimates; (e) furthered research and development for method of assessing the entropy of model residuals; (f) developed Bayesian framework for obtaining signal estimates from combined public and private information sources; (g) completed a corrigendum, extending subsampling results for Lipschitz continuous statistics; (h) furthered development and discussion of vector band pass and low pass filters; (i) completed extensive revisions and simulations for work on non-nested model comparisons; and (j) developed software to compute autocorrelations for a spatial long memory process.

Staff: Tucker McElroy (x33227), Brian Monsell, James Livsey, Osbert Pang, Anindya Roy

C. Time Series Model Development
Description: This work develops a flexible integer-valued autoregressive (AR) model for count data that contain data over- or under-dispersion (i.e. count data where the variance is larger or smaller than the mean, respectively). Such a model will contain Poisson and Bernoulli AR models as special cases.

Highlights: Staff continue to develop theoretical results and computational codes in R to analyze relevant data.

Staff: Kimberly Sellers (x39808)

Experimentation and Statistical Modeling

Motivation: Experiments at the Census Bureau are used to answer many research questions, especially those related to testing, evaluating, and advancing survey sampling methods. A properly designed experiment provides a valid, cost-effective framework that ensures the right type of data is collected as well as sufficient sample sizes and power are attained to address the questions of interest. The use of valid statistical models is vital to both the analysis of results from designed experiments and in characterizing relationships between variables in the vast data sources available to the Census Bureau. Statistical modeling is an essential component for wisely integrating data from previous sources (e.g., censuses, sample surveys, and administrative records) in order to maximize the information that they can provide.

Research Problems:

- Investigate bootstrap methodology for sample surveys; implement the bootstrap under complex sample survey designs; investigate variance estimation for linear and non-linear statistics and confidence interval computation; incorporate survey weights in the bootstrap; investigate imputation and the bootstrap under various non-response mechanisms.

- Investigate methodology for experimental designs embedded in sample surveys; investigation of large-scale field experiments embedded in ongoing surveys; design based and model based analysis and variance estimation incorporating the sampling design and the experimental design; factorial designs embedded in sample surveys and the estimation of interactions; testing non-response using embedded experiments. Use simulation studies.
- Assess feasibility of established design methods (e.g., factorial designs) in Census Bureau experimental tests.
- Identify and develop statistical models (e.g., loglinear models, mixture models, and mixed-effects models) to characterize relationships between variables measured in censuses, sample surveys, and administrative records.
- Assess the applicability of post hoc methods (e.g., multiple comparisons and tolerance intervals) with future designed experiments and when reviewing previous data analyses.

Potential Applications:

- Modeling approaches with administrative records can help enhance the information obtained from various sample surveys.
- Experimental design can help guide and validate testing procedures proposed for the 2020 Census.
- Expanding the collection of experimental design procedures currently utilized with the American Community Survey.

A. Design and Analysis of Embedded Experiments

Description: This ongoing project will explore rigorous analysis of embedded experiments: from simple idealized designs to complex designs used in practice at the Census Bureau.

Highlights: The Census Bureau regularly conducts large-scale field experiments embedded in surveys. For example, experiments are carried out within the American Community Survey (ACS) to test various strategies of contacting respondents. To obtain correct inference, statistical analysis of an embedded experiment must take both sample design and experimental design into account.

Staff: Thomas Mathew (x35337), Andrew Raim

B. Synthetic Survey and Processing Experiments

Description: To improve operational efficiencies and reduce costs of survey processing, this project will simulate a survey, in which an artificial team of interviewers seek out an artificial set of respondents, to test alternative methods of allocating resources in the field and to test alternatives for the post-processing of the gathered survey data. When calibrated with survey paradata, the model may also serve as a test bed for new methods of missing data imputation.

Highlights: This project is currently on hold.

Staff: TBA

C. Multivariate Nonparametric Tolerance Regions

Description: A tolerance region for a multivariate population is a region computed using a random sample that will contain a specified proportion or more of the population, with a given confidence level. Typically, tolerance regions that have been computed for multivariate populations are elliptical in shape. A difficulty with an elliptical region is that it cannot provide information on the individual components of the measurement vector. However, such information can be obtained if we compute tolerance regions that are rectangular in shape. This project applies bootstrap ideas to compute multivariate tolerance regions in a nonparametric framework. Such an approach can be applied to multivariate economic data and aid in the editing process by identifying multivariate observations that are outlying in one or more attributes and subsequently should undergo further review.

Highlights: No significant updates these quarters.

Staff: Thomas Mathew (x35337)

D. Master Address File (MAF) Research—Developing a Generalized Regression Model for Count Data

Description: This project develops a zero-inflated version of a generalized regression model for count data based on the Conway-Maxwell-Poisson distribution to allow for data-dispersion and excess zeroes in the dataset. The objective of this project is to develop and consider an alternative regression model for use to describe associations with changes in the number of housing units (adds or deletes) on a block, and predict where housing growth or decline may occur in the MAF.

Highlights: A manuscript describing the statistical methodology associated with this work was accepted for publication in *Computational Statistics & Data Analysis.*

Staff: Kimberly Sellers (x39808), Andrew Raim

E. Modeling the Causal Effects of Field Representative Actions and Strategies

Description: Field Representatives (FRs) apply different strategies for managing their monthly workloads. For example, some FRs may place high priority on contacting households that are perceived as likely to respond, putting aside the more difficult cases until later in the month. With large volumes of information flowing from paradata systems, we are better able to model FR data collection behavior. However, to understand the causal effects of these behaviors on outcomes of interest (response rates, measures of data quality and measures of cost), we need to adjust for confounding characteristics of FRs and their caseloads. In this project, we are developing techniques for causal inference from observational (non-experimental) data on FR characteristics, behaviors and performance measures.

Highlights: This project is currently on hold.

Staff: Doug Galagate, Robert Ashmead

F. Development of a Bivariate Distribution for Count Data where Data Dispersion is Present

Description: This project develops a bivariate form of the Conway-Maxwell-Poisson distribution to serve as a tool to describe variation and association for two count variables that express over- or under-dispersion (relationships where the variance of the data is larger or smaller than the mean, respectively).

Highlights: No significant updates these quarters.

Staff: Kimberly Sellers (x39808), Darcy Steeg Morris

G. Developing a Flexible Stochastic Process for Significantly Dispersed Count Data

Description: The Bernoulli and Poisson are two popular count processes; however, both rely on strict assumptions that motivate their use. CSRM staff (with other collaborators) instead propose a generalized count process (hereafter named the Conway-Maxwell-Poisson process) that not only includes the Bernoulli and Poisson processes as special cases, but also serves as a flexible mechanism to describe count processes that approximate data with over- or under-dispersion. Staff introduce the process and its associated generalized waiting time distribution with several real-data applications to illustrate its flexibility for a variety of data structures. This new generalized process will enable analysts to better model count processes where data dispersion exists in a more accommodating and flexible manner.

Highlights: Staff are revising a manuscript for publication.

Staff: Kimberly Sellers (x39808), Darcy Steeg Morris

Simulation and Statistical Modeling

Motivation: Simulation studies that are carefully designed under realistic survey conditions can be used to evaluate the quality of new statistical methodology for Census Bureau data. Furthermore, new computationally intensive statistical methodology is often beneficial because it can require less strict assumptions, offer more flexibility in sampling or modeling, accommodate complex features in the data, enable valid inference where other methods might fail, etc. Statistical modeling is at the core of the design of realistic simulation studies and the development of intensive computational statistical methods. Modeling also enables one to efficiently use all available information when producing estimates. Such studies can benefit from software such as *Tea* for data processing. Statistical disclosure avoidance methods are also developed and properties studied.

Research Problems:

- Systematically develop an environment for simulating complex surveys that can by used as a test-bed for new data analysis methods.
- Develop flexible model-based estimation methods for survey data.
- Develop new methods for statistical disclosure control that simultaneously protect confidential data from disclosure while enabling valid inferences to be drawn on relevant population parameters.
- Investigate the bootstrap for analyzing data from complex sample surveys.
- Continue to formalize the codebase and user interfacing for *Tea*, especially within the context of the current enterprise environment.
- Develop models for the analysis of measurement errors in Demographic sample surveys (e.g., Current Population Survey or the Survey of Income and Program Participation).
- Identify and develop statistical models (e.g., loglinear models, mixture models, and mixed-effects models) to characterize relationships between variables measured in censuses, sample surveys, and administrative records.
- Investigate noise multiplication for statistical disclosure control.

Potential Applications:

- Simulating data collection operations using Monte Carlo techniques can help the Census Bureau make more efficient changes.
- Use noise multiplication or synthetic data as an alternative to top coding for statistical disclosure control in publicly released data. Both noise multiplication and synthetic data have the potential to preserve more information in the released data over top coding.
- Rigorous statistical disclosure control methods allow for the release of new microdata products.
- *Tea* provides modeling and editing flexibility, especially with a focus on incorporating administrative data.
- Using an environment for simulating complex surveys, statistical properties of new methods for missing data imputation, model-based estimation, small area estimation, etc. can be evaluated.

- Model-based estimation procedures enable efficient use of auxiliary information (for example, Economic Census information in business surveys), and can be applied in situations where variables are highly skewed and sample sizes are not sufficiently large to justify normal approximations. These methods may also be applicable to analyze data arising from a mechanism other than random sampling.
- Variance estimates and confidence intervals in complex surveys can be obtained via the bootstrap.
- Modeling approaches with administrative records can help enhance the information obtained from various sample surveys.

A. Development and Evaluation of Methodology for Statistical Disclosure Control

Description: When survey organizations release data to the public, a major concern is the protection of individual records from disclosure while maintaining quality and utility of the released data. Procedures that deliberately alter data prior to their release fall under the general heading of statistical disclosure control. This project develops new methodology for statistical disclosure control, and evaluates properties of new and existing methods. We develop and study methods that yield valid statistical analyses, while simultaneously protecting individual records from disclosure.

Highlights: Staff continued work on the development of new and exact methods for drawing parametric inference based on singly imputed partially synthetic data generated via plug-in sampling. Staff developed this methodology for the cases when the original data follow either a multivariate normal or a multiple linear regression model. Staff defined sufficient conditions under which the methodology will yield valid inference, and studied properties of the methodology for the multiple linear regression model when certain conditions do not hold. Specifically, staff studied the scenario where the original data follow a linear regression model, and the data analyst observes a set of singly imputed synthetic data; however, the data generating model, imputation model, and data analysis model are not all the same. Our analysis includes both theoretical and empirical results to evaluate how inference is affected. Staff also studied another scenario where the sufficient conditions for valid inference do not hold because the statistical agency uses the regression of y on x to generate synthetic data, but the data analyst's model is the regression of x on y. Under each of these scenarios, staff compared the performance of our methodology for singly imputed synthetic data with the performance of established methods for multiply imputed synthetic data. Staff revised a manuscript entitled, "Likelihood Based Finite Sample Inference for Singly Imputed Synthetic Data Under the Multivariate Normal and Multiple Linear Regression Models." All of the material discussed above is included in the revision, and the manuscript was accepted for publication in the *Journal of Privacy and Confidentiality*.

Staff completed our manuscript entitled, "Inference for Multivariate Regression Model based on Synthetic Data generated using Plug-in Sampling," where staff develop likelihood based inference based on both singly and multiply imputed synthetic data, generated via plug-in sampling, under a multivariate linear regression model. Staff also completed our manuscript entitled, "Inference for Multivariate Regression Model based on Synthetic Data generated under Fixed-Posterior Predictive Sampling: Comparison with Plug-in Sampling," where staff develop likelihood based inference based on both singly and multiply imputed synthetic data, generated via fixed-posterior predictive sampling, under a multivariate linear regression model. In this manuscript, staff also provide some comparisons between fixed-posterior predictive sampling and plug-in sampling in terms of quality of inference and privacy protection.

Staff: Martin Klein (x37856), Bimal Sinha (CDAR), Thomas Mathew, Brett Moran

Summer at Census

Description: For each summer since 2009, recognized scholars in the following and related fields applicable to censuses and large-scale sample surveys are invited for short-term visits (one to five days) primarily between May and September: statistics, survey methodology, demography, economics, geography, social and behavioral sciences, and computer science. Scholars present a seminar based on their research and engage in collaborative research with Census Bureau researchers and staff.

Scholars are identified through an annual Census Bureau-wide solicitation by the Center for Statistical Research and Methodology.

Highlights: On March 7, 2016 staff sent out the Call for 2016 *SUMMER AT CENSUS* Nominations.

Staff: Tommy Wright (x31702), Michael Leibert

Research Support and Assistance

This staff provides substantive support in the conduct of research, research assistance, technical assistance, and secretarial support for the various research efforts.

Staff: Alisha Armas, Erica Magruder, Kelly Taylor

3. PUBLICATIONS

3.1 JOURNAL ARTICLES, PUBLICATIONS

Abramowitz, J., O'Hara, B., and Morris, D.S. (In Press). “Risking Life and Limb: Estimating a Measure of Medical Care Economic Risk and Considering its Implications,” *Health Economics*.

Blakely, C. and McElroy, T. (In Press). “Signal Extraction Goodness-of-fit Diagnostic Tests Under Model Parameter Uncertainty,” *Econometrics Reviews.*

Franco, C. and Bell, W. R. (2015). “Borrowing Information Over Time in Binomial/Logit Normal Models for Small Area Estimation,” Joint issue of *Statistics in Transition* and *Survey Methodology, 16 (4):* 563-584.

Holan, S., McElroy, T., and Wu, G. (In Press). “The Cepstral Model for Multivariate Time Series: The Vector Exponential Model,” *Statistica Sinica.*

Janicki, R. and McElroy, T. (2016). “Hermite Expansion and Estimation of Monotonic Transformations of Gaussian Data,” *Journal of Nonparametric Statistics, 28(1):* 207--234.

Klein, M. and Sinha, B. (2016). “Likelihood Based Finite Sample Inference for Singly Imputed Synthetic Data Under the Multivariate Normal and Multiple Linear Regression Models,” *Journal of Privacy and Confidentiality,7*: 43-98.

Lu, X. and West, D.B. (In Press). “New Proof that 4-connected Planar Graphs are Hamiltonian-connected,*”* *Discussiones Mathematicae Graph Theory*.

McElroy, T. (In Press). “Multivariate Seasonal Adjustment, Economic Identities, and Seasonal Taxonomy,” *Journal of Business and Economics Statistics.*

McElroy, T. and Holan, S. (2016). “Estimation of Time Series with Multiple Long-Range Persistencies,” *Computational Statistics and Data Analysis, 101*: 44--56.

McElroy, T. and McCracken, M. (Published online). “Multi-Step Ahead Forecasting of Vector Time Series,” *Econometrics Reviews.*

McElroy, T. and Nagaraja, C. (2016). “Tail Index Estimation with a Fixed Tuning Parameter Fraction,” *Journal of Statistical Planning and Inference, 170:* 27--45.

Morris, D.S., Keller, A., and Clark, B. (In Press). “An Approach for Using Administrative Records to Reduce Contacts in the 2020 Census," *Statistical Journal of the International Association for Official Statistics*.

Sellers, K.F., Morris, D.S., and Balakrishnan, N. (2015). “Bivariate Conway-Maxwell-Poisson Distribution: Formulation, Properties, and Inference,” *Journal of Multivariate Analysis* (*JMVA*-15-142; November 11, 2015).

Sellers, K.F. and Raim, A.M. (2016). “A Flexible Zero-inflated Model to Address Data Dispersion,” *Computational Statistics and Data Analysis, 99*: 68-80.

Trimbur, T. and McElroy, T. (In Press). “Signal Extraction for Nonstationary Time Series With Diverse Sampling Rules,” *Journal of Time Series Econometrics.*

Wildi, M. and McElroy, T. (In Press). “Optimal Real-Time Filters for Linear Prediction Problems,” *Journal of Time Series Econometrics.*

Young, D.S., Raim, A.M., and Johnson, N.R. (In Press). “Zero-inflated Modelling for Characterizing Coverage Errors of Extracts from the U.S. Census Bureau's Master Address File,” *Journal of the Royal Statistical Society: Series A.*

3.2 BOOKS/BOOK CHAPTERS

Christen, P. and Winkler, W. E. (To Appear). “Record Linkage,” in *Encyclopedia of Machine Learning and Data Mining*.

Winkler, W. E. (2015). “Probabilistic Linkage,” in Goldstein, H., Harron, K., and Dibbel, C. (Eds.), *Methodological Developments in Data Linkage*. Wiley.

3.3 PROCEEDINGS PAPERS

FCSM Proceedings, Federal Committee on Statistical Methodology Meeting, Washington, D.C., December 1-3, 2015.

- Laura Bechtel, Darcy Steeg Morris, and Katherine Jenny Thompson, “Using Classification Trees to Recommend Hot Deck Imputation Methods: A Case Study”.

Joint Statistical Meetings, American Statistical Association, Seattle, Washington, August 8-13, 2015.
2015 Proceedings of the American Statistical Association

- Maria M. Garcia, Darcy Steeg Morris, and L. Kaili Diamond, “Implementation of Ratio Imputation and Sequential Regression Multivariate Imputation on Economic Census Products”, 1056-1070.
- Martin Klein, Joanna Fane Lineback, and Joseph L. Schafer, “Evaluating Imputation and Estimation Procedures in a Survey of Wholesale Businesses”, 1997-2008.
- Darcy Steeg Morris, Andrew Keller, and Brian Clark, “An Approach for Using Administrative Records to Reduce Contacts in the 2020 Census”, 3278-3292.
- Mary Mulry and Andrew Keller, “Are Proxy Responses Better Than Administrative Records?” 2465-2479.
- Andrew M. Raim, Marissa N. Gargano, Nagaraj K. Neerchal, and Jorge G. Morel, “Bayesian Analysis of Overdispersed Binomial Data Using Mixture Link Regression”, 2794-2808.

3.4 CENTER FOR STATISTICAL RESEARCH & METHODOLOGY RESEARCH REPORTS

<http://www.census.gov/srd/www/byyear.html>

RR (Statistics #2015-04): Andrew Raim and Marissa N. Gargano. “Selection of Predictors to Model Coverage Errors in the Master Address File,” December 30, 2015.

RR (Statistics #2016-01): Ryan Janicki. “Estimation of the Difference of Small Area Means from Different Time Periods,” February 25, 2016.

RR (Statistics #2016-02): Osbert Pang and Brian C. Monsell. “Examining Diagnostics for Trading-Day Effects from X-13ARIMA-SEATS,” March 11, 2016.

3.5 OTHER REPORTS

Hughes, T., Slud, E., Ashmead, R., Walsh, R. (2016). “Results of a Field Pilot to Reduce Respondent Contact Burden in the American Community Survey’s Computer Assisted Personal Interviewing Operation,” *American Community Survey Research and Evaluation Report Memorandum Series, ACS16-RER-07.*
http://www.census.gov/library/working-papers/2016/acs/2016_Hughes_01.html

4. TALKS AND PRESENTATIONS

Statistics Colloquium, University of Maryland, Baltimore County, Baltimore, Maryland, October 2, 2015.

- Robert Ashmead, "Propensity Score Estimators for Causal Inference with Complex Survey Data."

2015 Morehouse Mathematics Fair, Morehouse College, Atlanta, Georgia, October 8, 2015.

- Kimberly Sellers, "Don't Count on Poisson: Introducing a Flexible Alternative Distribution to Model Count Data."

Nielsen Office, Columbia, Maryland, *October 15, 2015.*

- Carolina Franco and William R. Bell, "Borrowing Information Over Time in Binomial/Logit Normal Models for Small Area Estimation."

Minisymposium Honoring Dianne O'Leary, SIAM Conference on Applied Linear Algebra, Atlanta, Georgia, October 26-30, 2015.

- Kimberly Sellers, "A Flexible Regression Model for Count Data."

Department of Statistics, University of Missouri, Columbia, Missouri, October 28, 2015.

- Martin Klein, "Noise Multiplication for Statistical Disclosure Control of Extreme Values in Log-normal Regression Samples."

NISS Workshop on Nonignorable Nonresponse, Bureau of Labor Statistics, Washington, D.C., November 12-13, 2015.

- Eric Slud, "Weighted Estimating Equations Based on Response Propensities in Terms of Covariates that are Observed only for Responders."

Workshop on Data Integration and Applications at the IEEE International Conference on Data Mining, Atlantic City, New Jersey, November 14, 2015.

- William E. Winkler, "Keynote: Clean-up and Preliminary Analysis for Data Mining Sets of National Files."

Time Series Workshop, Office of National Statistics, London, United Kingdom, November 19-20, 2015.

- Brian Monsell and James Livsey, "Overview of Time Series Issues and Research at the U.S. Census Bureau
- Brian Monsell, "Weekly Seasonal Adjustment."
- James Livsey, "Diagnostics for Deciding on Moving Holiday Window."

Federal Committee on Statistical Methodology Research Conference, Washington, D.C., December 1, 2015.

- Carolina Franco, Roderick J. Little, Thomas A. Louis, and Eric V. Slud, "Comparative Study of Confidence Intervals for Proportions in Complex Surveys."

Departmental Seminar, Department of Statistics, University of Kentucky, Lexington, Kentucky, December 4, 2015.

- Kimberly Sellers, "A Flexible Regression Model for Count Data."

Computational and Financial Econometrics, London, United Kingdom, December 12-14, 2015.

- Tucker McElroy, "Seasonal Adjustment of Meager Time Series."

8th International Conference of the ERCIM Working Group on Computational and Methodological Statistics (CMStatistics 2015), University of London, London, United Kingdom. December 13, 2015.

- Carolina Franco, Serena Arima, William R. Bell, Gauri Datta, and Brunero Liseo, "Bayesian Treatment of a Multivariate Fay-Herriot Functional Measurement Error Model with Applications."

Universidad Carlos III de Madrid, Madrid, Spain, December 19, 2015

- Carolina Franco and William R. Bell, "Temporal Extensions to a Hierarchical Model for Proportions from Complex Survey Data. Statistical Seminar."

Joint Program in Survey Methodology, University of Maryland, February 5, 2016

- William E. Winkler, “Clean-up and Preliminary Analysis of Sets of National Files.”

Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, Maryland, February 19, 2016.

- Andrew Raim, “An Extension of Generalized Linear Models to Finite Mixture Outcomes.”

Department of Statistics Colloquium, University of Kentucky, Lexington, Kentucky, February 19, 2016.

- Tommy Wright, “The Equivalence of Neyman Optimum Allocation for Sampling and Equal Proportions for Apportioning the U.S. House of Representatives.”

2016 Ross-Royall Symposium. Johns Hopkins University, Baltimore, Maryland. February 26, 2016.

- William R. Bell and Carolina Franco, “Combining Estimates from Related Surveys via Bivariate Models.”

Statistics Colloquium, University of Maryland, College Park, Maryland, March 4, 2016.

- Emanuel Ben-David, "Gaussian DAG models with Symmetries"

24th Symposium on Nonlinear Dynamics and Econometrics, University of Alabama, Tuscaloosa, Alabama, March 11, 2016.

- Thomas Trimbur (with Bill Bell), “The Effects of Seasonal Heterskedasticity in Time Series on Trend Estimation and Seasonal Adjustment.”

Cameroon International Conference on Recent Development in Applied Statistics, Yaounde, Cameroon, March 14-18, 2016.

- Tommy Wright (Keynote Address), “No Calculations When Observations Can Be Made.”

17th Annual OxMetrics User Conference, The George Washington University, Washington, D.C., March 18, 2016.

- Thomas Trimbur (with Bill Bell), “The Effects of Seasonal Heterskedasticity in Time Series on Trend Estimation and Seasonal Adjustment.”

Statistics Canada Methodology Symposium, Gatineau, Quebec, Canada, March 22 – 24, 2016.

- Mary Mulry, Elizabeth M. Nichols, and Jennifer Hunter Childs, “Using Administrative Records to Evaluate Survey Data.”

Business Week, University of Texas at Arlington, Arlington, Texas, March 28 - April 1, 2016.

- Mary Mulry. “Statistical Methods Used in Planning and Implementing the 2010 Census Communications Campaign.”

5. CENTER FOR STATISTICAL RESEARCH AND METHODOLOGY SEMINAR SERIES

William Winkler, U.S. Census Bureau, “Edit/Imputation Course,” October 1, 2015.

William Winkler and Edward Porter, U.S. Census Bureau, “Record Linkage Course,” October 20 & 21, 2015.

William Winkler, U.S. Census Bureau, “Quality and Analysis or Sets of National Files,” November 3, 2015.

Jared Murray, Carnegie Mellon University, “Multiple Imputation of Missing Categorical and Continuous Values via Bayesian Mixture Models with Local Dependence,” November 19, 2015.

Bret Hanlon, University of Wisconsin-Madison, “Robust Estimation for a Supercritical Branching Processor under Family-Size Sampling,” December 8, 2015.

Jason Bernstein, The Pennsylvania State University, “Time Series Analysis of Motor Proteins,” January 12, 2016.

Zachary Seeskin, Northwestern University, “Effects of Census Accuracy on Apportionment of Congress and Allocations of Federal Funds,” January 21, 2016.

Emanuel Ben-David, U.S. Census Bureau, “An Introduction to Probabilistic Graphical Models,” March 29, 2016.

6. PERSONNEL ITEMS

6.1 HONORS/AWARDS/SPECIAL RECOGNITION

6.2 SIGNIFICANT SERVICE TO PROFESSION

Robert Ashmead

- Refereed papers for *Journal of Survey Statistics and Methodology* and *The American Statistician*

Emanuel Ben-David

- Refereed papers for the 33rd International Conference on Machine Learning (ICML 2016), 19th International Conference on Artificial Intelligence and Statistics (AISTAT 2016), *Annals of Applied Statistics*, *Mathematical Reviews*, *Journal of Statistical Planning and Inference*, and *The American Statistician*.

Carolina Franco

- Refereed papers for *The American Statistician*

Martin Klein

- Refereed papers for *Journal of the International Association for Official Statistics* and *Statistical Papers*
- Member, Ph.D. Dissertation in Statistics Committee, University of Maryland, Baltimore County

Thomas Mathew

- Associate Editor, *Journal of the American Statistical Association*
- Associate Editor, *Statistical Methodology*
- Associate Editor, *Sankhya, Series B*
- Editorial Board, Member, *Journal of Occupational and Environmental Hygiene*
- Member, American Statistical Association's Committee on W.J. Youden Award in Inter-laboratory Testing

Tucker McElroy

- Refereed papers for *Annals of Statistics, Journal of Applied Econometrics, Journal of Official Statistics,* and *Communications in Statistics*

Mary Mulry

- Associate Editor, *Journal of Official Statistics*
- Methodology co-Editor, *Statistical Journal of the International Association of Official Statistics*
- Invited Session Organizer, Fifth International Conference on Establishment Surveys (ICES-V)

Kimberly Sellers

- Member, American Statistical Association Committee on Women in Statistics
- Associate Editor, *The American Statistician*
- Advisory Board Member and Director, BDN STEMers for International Black Doctoral Network Association, Incorporated
- Refereed papers for *Applied Stochastic Models in Business and Industry, Biometrics, Communications in Statistics – Theory and Methods, Computers & Industrial Engineering, Lifetime Data Analysis, Quality and Reliability Engineering International*, and *Statistics*
- Member, Scientific Program Committee, International Conference on Statistical Distributions and Applications (ICOSDA) 2016

Eric Slud

- Associate Editor, *Biometrika*
- Associate Editor, *Journal of Survey Statistics and Methodology*
- Associate Editor, *Lifetime of Data Analysis*

William Winkler

- Refereed papers for the *Journal of Official Statistics* and the *Journal of Survey Statistics and Methodology*
- Reviewed grant proposal and made recommendation related to large grant proposal by the Dutch government
- Associate Editor, *Journal of Privacy and Confidentiality*
- Associate Editor, *Transactions on Data Privacy*
- Member, Program Committee for *Statistical Data Protection 2016*
- Member, Program Committee for *IEEE 2015 ICDM Data Integration and Applications*
- Member, Program Committee for ACM Workshop on Population Informatics at KDD'16

Tommy Wright

- Associate Editor, *The American Statistician*
- Chair, Waksberg Award Committee, *Survey Methodology*
- Member, Board of Trustees, National Institute of Statistical Sciences

6.3 PERSONNEL NOTES

Alisha Armas completed graduate studies at American University and accepted another position.

Dan Weinberg (new Ph.D., Mathematics, University of Maryland, College Park) joined our Time Series Research Group.